StyleGAN-based Brain MRI Anomaly Detection via Latent Code Retrieval and Partial Swap

Jie Wei[1], Xiaofei Hu[2], Shaoting Zhang[1,3], and Guotai Wang[1,3]

[1] University of Electronic Science and Technology of China, Chengdu, 611731, China
[2] Department of Nuclear Medicine, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing, 400038, China
[3] Shanghai Artificial Intelligence Laboratory, Shanghai, 200030, China
guotai.wang@uestc.edu.cn

Abstract. Medical anomaly detection aims at identifying samples that deviate from normal patterns and localizing specific anomalous regions, playing a critical role in early detection and intervention of diseases. Reconstruction methods based on generative models are a key category among current methods for medical anomaly detection. However, a common challenge for them is achieving accurate reconstruction of normal regions while suppressing the reconstruction of anomalous regions. StyleGAN, with its powerful generative capability and the ability to perform controllable image modifications, has shown huge potential for medical image anomaly detection. However, the latent space of StyleGAN still requires further exploration and utilization. In this paper, we propose a StyleGAN-based latent Code Retrieval and Partial Swap (SCRPS) method for brain image anomaly detection. We construct a healthy image latent code repository by leveraging GAN inversion in StyleGAN's latent space. We then design a coarse-to-fine latent code retrieval mechanism to filter out normal images most similar to test image. We also introduce a partial latent code swap strategy that replaces anomalous latent codes with linear combinations of normal latent codes and employ a perceptual score to perform anomaly localization. Comprehensive experiments on brain tumor and stroke lesion datasets show that our method outperforms several state-of-the-art approaches, with 3.12 and 7.14 percentage points improvements in average volume-level AUROC and maximum achievable Dice score, respectively.

Keywords: Anomaly detection · Brain MRI · Generative models

1 Introduction

Anomaly detection, as a special case of out-of-distribution (OOD) detection [26], is crucial in early disease detection from brain Magnetic Resonance Images (MRI) by identifying samples that deviate from normal patterns and localizing specific anomalous regions, aiding doctors in decision-making and enhancing their efficiency [21]. Due to the high annotation cost of supervised training and

its inability to cover unseen anomalies [7], training on normal samples is more appealing and practical. Anomaly detection, which relies solely on normal samples for training, follows an unsupervised paradigm and is not limited by annotated anomaly types, thus significantly reducing the annotation burden on clinicians and proving especially valuable for detecting rare or novel lesions [6].

Current research on anomaly detection can generally be categorized into three main categories. Pseudo-outlier augmentation methods [29,33] synthesize anomalies and their corresponding labels on normal images, thereby converting the unsupervised task into supervised one. However, due to the diverse and unpredictable nature of anomalies, these methods cannot comprehensively simulate all potential anomaly types. Feature-based methods [8,24] typically detect anomalies by projecting images into a feature space using encoders pre-trained on natural images. With the emergence of large models, some methods [16,32] also leverage these models as feature extractors. However, such extractors may not be appropriately adapted for medical anomaly detection.

Reconstruction-based methods [2,12,25,28] assume that networks trained on normal data cannot effectively reconstruct anomalous images. Anomalies can therefore be detected through reconstruction errors. However, if anomalies share similar compositional patterns with normal data, or if the decoder has strong reconstruction capability, the anomalies may still be accurately reconstructed [30]. Therefore, suppressing the reconstruction of anomalous regions remains a key challenge for reconstruction-based methods. To deal with this problem, approaches like MemAE [12] and SQUID [28] enhance reconstruction networks with memory modules, replacing anomalous image features with weighted combinations of normal features stored in memory. Due to the online updating characteristics of the memory modules and their correspondence to pixel-level features, they perform well in anatomically aligned scenarios, such as chest X-rays. However, when applied to scenarios with diverse patterns, such as brain slices that are more difficult to be well aligned due to large variation of size, pose and and deformation caused by lesions, they tend to experience performance degradation or even complete failure.

Among reconstruction-based methods, using a Generative Adversarial Network (GAN) with an encoder-based GAN inversion mechanism to learn the mapping from pixel space to latent space for reconstruction is a typical approach, as seen in methods like f-AnoGAN [25] and Ganomaly [2]. However, traditional GANs lacks effective latent space decoupling properties, and the encoder-based GAN inversion method does not always guarantee high-quality reconstruction performance [27]. In contrast, StyleGAN [17] stands out for generation capability and ability to perform controllable image modifications through latent space. Combined with optimization-based GAN inversion, which treats latent space embeddings as learnable variables, StyleGAN can achieve excellent reconstruction result [1] and has been applied to medical image analysis tasks [14,15]. However, to fully leverage its potential in anomaly detection, the latent space of StyleGAN still requires further exploration and utilization.

In this work, we propose a novel latent code retrieval and partial swap method based on StyleGAN reconstruction for brain MRI anomaly detection. Our contributions are summarized as follows: 1) We explore the latent space of StyleGAN for anomaly detection and build an off-line normal image latent code repository using GAN inversion; 2) We design a coarse-to-fine retrieval strategy to filter normal images similar to test images from the normal image latent code repository to suppress the reconstruction of anomalous regions. 3) We propose a partial latent code swap strategy to ensure accurate reconstruction of normal regions for reducing false positives.

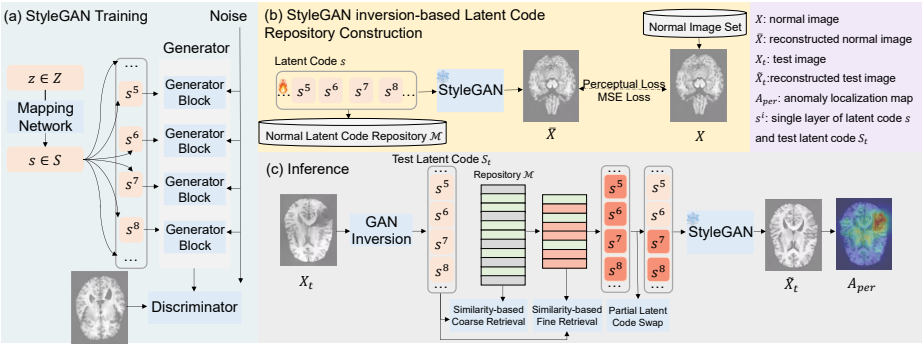

Fig. 1. Overview of the proposed method. (a) The training process of StyleGAN on normal images. (b) Normal latent code repository construction based on GAN inversion. (c) Coarse-to-fine normal latent code retrieval and partial latent code swap strategy to reconstruct corresponding normal image.

2 Method

An overview of our method is presented in Fig. 1. First, a StyleGAN model is trained on normal images of healthy individuals. Second, GAN inversion is performed in the latent space S of each training image, and the resulting latent codes are stored as a normal latent code repository $\mathcal{M}$. During inference, a test image X_t is projected into latent space through GAN inversion to obtain its latent code s_t. Coarse-to-fine retrieval and combination of normal latent codes are performed to obtain a fused normal latent code $\widehat{s}$, which undergoes a partial swap operation with s_t before being input into StyleGAN to obtain a reconstructed image $\tilde{X}_t$. Finally, perceptual score is used to localize anomalous regions.

2.1 StyleGAN Inversion-based Latent Code Repository Construction

Overview of StyleGAN. StyleGAN [17] is a generative adversarial network that consists of a generator $G(\cdot)$ and a discriminator $D(\cdot)$. As shown in Fig. 1(a), a random noise vector z is projected to a high-dimensional latent space $S \in \mathbb{R}^{L\times 512}$ (W^+ space mentioned in [1]) through a mapping network. The generator $G(\cdot)$ with L layers accepts latent code s and noise b as input to generate image. $G(\cdot)$ generates high-resolution images progressively from low-resolution ones, where s controls the content of each intermediate layer of the generator. The shallow-layer code primarily captures structural and global features, while the deeper-layer code focuses on fine-grained details. The discriminator [13] differentiates between generated images and real images, providing supervision and driving the optimization of the entire network.

Latent Code Generation for Training Images. GAN inversion [27] aims at mapping an input image from pixel space into latent space, where various real image editing tasks can be efficiently performed [23,10]. The key challenge of GAN inversion is to find the appropriate inversion space that minimizes distortion while maintaining editability [19]. Existing research [19] demonstrates that for StyleGAN, the W^+ space can strike a balance between distortion and editability. Therefore, we conduct GAN inversion in that latent space and build a latent feature repository using normal images.

Specifically, we keep noise input b fixed for all the training images, and follow a straightforward optimization process to optimize the latent code s for each image via a loss function combining perceptual loss [31] and MSE loss between input image X and reconstructed image $\overline{X}$.

$$s(X) = \min_{s} L_{percept}(X, \overline{X}) + \frac{\lambda_{mse}}{N}\|X - \overline{X}\|_2^2 \tag{1}$$

where $\overline{X} = G(s, b)$, N is the number of pixles in the image, and λ_{mse} is a weight factor.

$$L_{percept}(X, \overline{X}) = \sum_{j=1}^{5} \frac{\|F_j(X) - F_j(\overline{X})\|_2^2}{N_j} \tag{2}$$

where F_j is the feature output of VGG16 [31] after the j-th convolutional block. N_j is the number of pixels in the j-th layer output. We optimize each normal image from the training set to obtain the latent code repository $\mathcal{M} \in \mathbb{R}^{K\times L\times 512}$, where K is the size of dateset.

2.2 Latent Code Retrieval, Combination and Partial Swap

Coarse-to-Fine Latent Code Retrieval and Combination. During the inference stage, we first apply GAN inversion to the test image X_t to derive its

corresponding latent code s_t based on Eq. 1. Next, we retrieve the most similar normal latent code from the latent code repository. However, given that the shallow and deep parts of latent codes exert relatively decoupled control over image content [34], directly comparing the entire latent code fails to explicitly capture this inherent relationship while suffering from ineffective similarity computation in high-dimensional spaces, we implement a coarse-to-fine latent code retrieval strategy as follows:

1. Coarse latent code matching: We initially preserve the first L_1 layers from both test latent code s_t and latent code repository $\mathcal{M}$, denoted as $s_t^< = s_t[0 : L_1 - 1]$ and $\mathcal{M}^< = \{s[0 : L_1 - 1] \text{ for } s \in \mathcal{M}\}$. We then compute the cosine similarity between $s_t^<$ and each latent code in $\mathcal{M}^<$, resulting in a subset $\mathcal{P}$ that contains top K_1 candidate latent codes with global shapes and distributions similar to test image X_t.
$$\mathcal{P} = \left\{ s \mid s^< \in \mathcal{M}^<, sim(s_t^<, s^<) > sim_{K_1}(s_t^<, \mathcal{M}^<) \right\} \tag{3}$$
where $sim_{K_1}(s_t^<, \mathcal{M}^<)$ denotes the K_1-th highest value for similarity between $s_t^<$ and items in $\mathcal{M}^<$.
2. Fine-grained detail refinement: We subsequently preserve the residual layers from both test latent code s_t and $\mathcal{P}$ to establish fine-grained detail alignment, let $s_t^> = s_t[L_1 : L-1]$ and $\mathcal{P}^> = \{s[L_1 : L-1] \text{ for } s \in \mathcal{P}\}$ denote the deeper part of the latent codes, we then filter top $K_2 (K_2 < K_1)$ samples with the most similar details to that in X_t.
$$\mathcal{P}' = \left\{ s \mid s^> \in \mathcal{P}^>, sim(s_t^>, s^>) > sim_{K_2}(s_t^>, \mathcal{P}^>) \right\} \tag{4}$$
where $sim_{K_2}(s_t^>, \mathcal{P}^>)$ denotes the K_2-th highest value for similarity between $s_t^>$ and items in $\mathcal{P}^>$.

After obtaining the selected subset $\mathcal{P}'$, we combine its items to obtain an aggregated latent code $\widehat{s} = \sum_{i=1}^{K_2} w_i s_i$, where $w_i = \frac{sim(s_t{}^>, s_i{}^>)}{\sum_{i=1}^{K_2} sim(s_t{}^>, s_i{}^>)}$ for $s_i \in \mathcal{P}'$.

Partial Latent Code Swap. Although the latent code $\widehat{s}$ can generate a normal image resembling the test image through the generator network, the distribution shift between $\mathcal{M}$ and s_t and distortion caused by anomalous lesions to normal regions often lead to discrepancies in the details of the normal regions between the synthesized image and the test image. As a result, the synthesized normal regions may lack precise alignment with the healthy patterns in the test image, resulting in false positives. In anomalous images, the shallow latent code control the overall image shape and appearance, and the local anomalies which manifests more in texture, density, and local morphology is more controlled by the deeper latent code. Therefore, we propose to partially swap s_t and $\widehat{s}$, aiming to preserving the normal areas of the test image as much as possible where suppressing the reconstruction of anomalous regions. The swap process is as follows:
$$\tilde{s} = s_t[0 : L_2 - 1] \oplus \widehat{s}[L_2 : L - 1] \tag{5}$$
where $\tilde{s}$ is the swapped latent code, L_2 is the swap starting layer, $\oplus$ is the concatenation operation.

Anomaly Localization based on Perceptual Score. During the anomaly localization phase, we input $\tilde{s}$ into the pre-trained StyleGAN generator $G(\cdot)$ to obtain reconstructed image $\tilde{X}_t = G(\tilde{s}, b)$. Subsequently, we employ the normalized perceptual score A_{per} proposed in [31] to localize the anomaly regions to better match human eye perception. Similar to the perceptual loss in Section 2.1, perceptual score uses a pre-trained VGG to extract feature maps from the residual map between X_t and $\tilde{X}_t$, and upsamples them to the original image size. A_{per} score for each pixel means the likelihood of being anomaly.

3 Experiments and Results

3.1 Datasets and Implementation Details

In this work, we focus on unsupervised anomaly localization in brain MRI images. For training, we selected 640 T1-weighted MRI scans of healthy individuals from the OpenBHB dataset [9], which contains over 5,000 healthy brain scans. For testing, we used two datasets: the BraTS2020 dataset [20,3,4], consisting of 369 brain MRI scans for tumors (only T1-weighted images used in our experiment), and the ATLAS dataset [18], consisting of 655 T1-weighted MRI scans of ischemic stroke lesions. Since StyleGAN is 2D-based, we focused on axial slices with lesions, and resized them to 160×128 pixels. For the ATLAS dataset, skull stripping was performed using FreeSurfer [11], and we conducted histogram equalization, histogram matching, and normalization for all data.

We trained StyleGAN V2 [17] with a learning rate of 0.002 for 12,000 iterations. To increase data diversity, we applied two sagittal-axis rotations (ranging from -15° to 15°) during training. We set the noise input b of StyleGAN's generator $G(\cdot)$ to a fixed value zero to focus only on the effect of the latent code on image generation. For efficiency consideration, we randomly selected 5% of the training data for GAN inversion, performing 2,000 iterations per slice with a learning rate of 0.01. The MSE loss weight λ_{mse} was 10, and the latent code dimension was 12×512. The hyper-parameter setting was $K_1 = 100, K_2 = 20, L_1 = 6$ and $L_2 = 9$. For evaluation metrics, we measured volume-level Area Under the ROC Curve (AUROC) and maximum achievable Dice coefficient ($\lceil Dice \rceil$) results as [5,7].

3.2 Comparison with State-of-the-art Methods

We compared our method with eight state-of-the-art unsupervised anomaly detection methods including reconstruction-based f-AnoGAN [25], Ganomaly [2], and MemAE [12]; feature-based PatchCore [24] and PaDiM [8]; hybrid methods DRÆM [29] and RealNet [33] that combine pseudo-outlier augmentation with reconstruction; and hybrid method MediCLIP [32] that combine pseudo-outlier augmentation with feature comparision. To ensure fair comparison, we used perceptual score A_{per} [31] for all the reconstruction-based methods for anomaly localization. For methods that require feature extraction such as PatchCore [24],

Table 1. Anomaly localization performance of different methods. $*$ denotes significant improvement from the best existing method with p-value < 0.05, assessed using a paired t-test.

Method	BraTS2020		ATLAS	
	AUROC	$\lceil DICE \rceil$	AUROC	$\lceil DICE \rceil$
f-AnoGAN [25]	83.52±7.31	24.42±11.72	88.17±5.25	18.09±9.57
Ganomaly [2]	85.41±5.42	26.83±11.60	86.01±5.71	15.60±8.88
MemAE [12]	85.30±6.53	26.26±11.23	84.47±6.04	13.94±9.15
PaDiM [8]	81.15±6.69	22.57±10.98	84.14±6.83	15.36±9.48
PatchCore [24]	80.63±6.76	20.08±10.00	87.76±6.11	19.07±9.93
DRÆM [29]	82.11±5.74	21.09±9.78	66.68±10.99	6.36±7.57
MediCLIP [32]	82.39±10.41	23.35±14.27	87.64±6.53	15.44±10.19
RealNet [33]	80.98±7.34	20.51±10.64	83.54±5.67	13.88±9.38
Ours	**89.43±7.03***	**37.04±14.54***	**90.38±4.97***	**24.14±13.05***

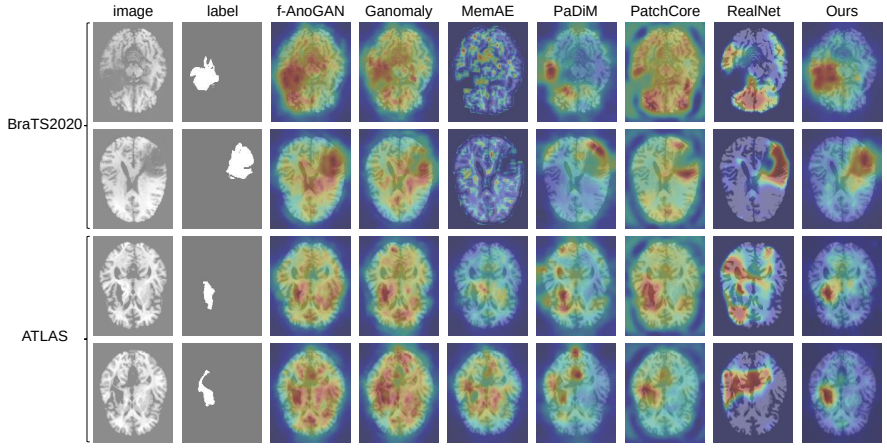

Fig. 2. Visual comparison for anomaly localization performance of different methods. Heater color indicates higher probability of being anomaly.

PaDiM [8], RealNet [33] and perceptual score calculation, we used feature maps at the last three convolutional blocks in VGG16 [31]. MediCLIP [32] used the image encoder of CLIP [22] as feature extractor.

Quantitative evaluation results of these methods on the two datasets are shown in Table 1. Our method achieves superior performance across all evaluation metrics on both datasets. On the BraTS2020 dataset, compared with the best existing method Ganomaly [2], our approach demonstrates improvements of 4.02 and 9.21 percentage points in terms of volume-level AUROC and volume-level $\lceil DICE \rceil$, respectively. On the ATLAS dataset, our method outperforms the best existing method f-AnoGAN [25] on AUROC by 2.21 percentage points and best existing method PatchCore [24] on $\lceil DICE \rceil$ by 5.07 percentage points, respectively. Fig. 2 presents qualitative visualizations of anomaly localization

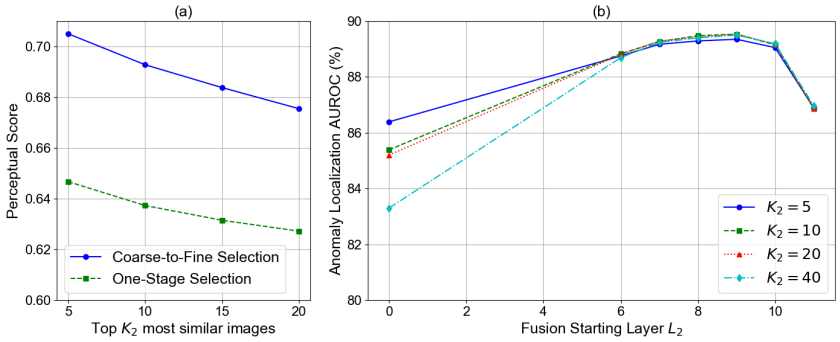

Fig. 3. Effect of hyper-parameters on BraTS2020 test set. (a) shows the average perceptual scores between test image and retrieved images with different K_2 values, (b) shows the impact of K_2 and swap starting layer L_2 on volume-level AUROC.

results from these methods. Our method demonstrates more precise anomaly localization ability with reduced false positive than the other methods.

3.3 Ablation Studies

Effectiveness of Coarse-to-fine Latent Code Retrieval. To validate the effectiveness of our coarse-to-fine latent code retrieval strategy, we conducted ablation studies on the BraTS2020 test set with two approaches: 1) One-stage latent code retrieval that selects K_2 latent codes from $\mathcal{M}$ based on $sim(s_t, s)$ for $s \in \mathcal{M}$ and 2) Our coarse-to-fine method with $L_1 = 6$. Fig. 3 (a) shows the average perceptual scores between X_t and the top K_2 images retrieved by two approaches when K_2 ranges from 5 to 20. Our proposed coarse-to-fine retrieval strategy consistently achieves higher perceptual score, proving its effectiveness in retrieving similar images. We further explored the impact and K_2 on the experimental results and set $K_1 = 5K_2$. As shown in Fig. 3 (b), once the normal latent code repository is fully constructed and latent code swap is applied with an appropriate L_2 value (e.g., 6-10), our method is not sensitive to K_2.

Effectiveness of Partial Latent Code Swap. We further investigated the impact of swap starting layer L_2 on localization performance, as shown in Fig. 3 (b). $L_2 = 0$ means not using latent code swap, and the volume-level AUROC was 4% lower than the best-performing latent code swap strategy. As we progressively replaced from deeper layer within the latent code, the performance of anomaly detection firstly improved and then declined. The AUROC and $\lceil DICE \rceil$ scores achieved their highest values when $L_2 = 9$. This means that a balance is achieved between preserving normal regions and replacing anomalous regions.

4 Conclusion

In conclusion, we proposed SCRPS, a StyleGAN-based latent code retrieval and partial swap method for brain MRI anomaly detection. By leveraging StyleGAN's reconstruction ability and latent space editability, we built a normal latent code repository for healthy images. During inference, a coarse-to-fine latent code retrieval strategy is used to obtain the most similar latent codes to reconstruct a normal image resembling the test image. To minimize false positives, we designed a partial latent code swap approach to retain the global structure of the test image. Experimental results on brain tumor and stroke image datasets showed the effectiveness of our method. In the future, it is of interest to explore the impact of our method on other types of anomaly detection.

Acknowledgement. This work was supported by the National Natural Science Foundation of China under grant 62271115 and Senior Medical Talents Program of Chongqing for Young and Middle-aged, Natural Science Foundation of Chongqing (cstc2021jcyj-msxmX0148).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Abdal, R., Qin, Y., Wonka, P.: Image2StyleGAN: How to embed images into the StyleGAN latent space? In: ICCV. pp. 4432–4441 (2019)
2. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: ACCV. pp. 622–637 (2019)
3. Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J.B., Farahani, K., Davatzikos, C.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific Data **4**(1), 1–13 (2017)
4. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., et al.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv preprint arXiv:1811.02629 (2018)
5. Behrendt, F., Bhattacharya, D., Mieling, R., Maack, L., Krüger, J., Opfer, R., Schlaefer, A.: Leveraging the mahalanobis distance to enhance unsupervised brain MRI anomaly detection. In: MICCAI. pp. 394–404 (2024)
6. Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A.: Reversing the abnormal: Pseudo-healthy generative networks for anomaly detection. In: MICCAI. pp. 293–303 (2023)
7. Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A.: Diffusion models with implicit guidance for medical anomaly detection. In: MICCAI. pp. 211–220 (2024)
8. Defard, T., Setkov, A., Loesch, A., Audigier, R.: PaDiM: A patch distribution modeling framework for anomaly detection and localization. In: ICPR. pp. 475–489 (2021)

9. Dufumier, B., Grigis, A., Victor, J., Ambroise, C., Frouin, V., Duchesnay, E.: OpenBHB: a large-scale multi-site brain MRI data-set for age prediction and debiasing. NeuroImage **263**, 119637 (2022)
10. Feng, X., Lin, J., Feng, C.M., Lu, G.: GAN inversion-based semi-supervised learning for medical image segmentation. Biomedical Signal Processing and Control **88**, 105536 (2024)
11. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
12. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: ICCV. pp. 1705–1714 (2019)
13. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
14. Hagag, A., Gomaa, A., Kornek, D., Maier, A., Fietkau, R., Bert, C., Huang, Y., Putz, F.: Deep learning for cancer prognosis prediction using portrait photos by StyleGAN embedding. In: MICCAI. pp. 198–208 (2024)
15. Hochberg, D.C., Greenspan, H., Giryes, R.: A self supervised StyleGAN for image annotation and classification with extremely limited labels. IEEE Transactions on Medical Imaging **41**(12), 3509–3519 (2022)
16. Huang, C., Jiang, A., Feng, J., Zhang, Y., Wang, X., Wang, Y.: Adapting visual-language models for generalizable anomaly detection in medical images. In: CVPR. pp. 11375–11385 (2024)
17. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: CVPR. pp. 8110–8119 (2020)
18. Liew, S.L., Lo, B.P., Donnelly, M.R., Zavaliangos-Petropulu, A., Jeong, J.N., Barisano, G., Hutton, A., Simon, J.P., Juliano, J.M., Suri, A., et al.: A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. Scientific Data **9**(1), 320 (2022)
19. Liu, H., Song, Y., Chen, Q.: Delving StyleGAN inversion for image editing: A foundation latent space viewpoint. In: CVPR. pp. 10072–10082 (2023)
20. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Transactions on Medical Imaging **34**(10), 1993–2024 (2014)
21. Pang, G., Shen, C., Cao, L., Hengel, A.V.D.: Deep learning for anomaly detection: A review. ACM Computing Surveys (CSUR) **54**(2), 1–38 (2021)
22. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: ICML. pp. 8748–8763 (2021)
23. Ren, Z., Stella, X.Y., Whitney, D.: Controllable medical image generation via GAN. Journal of Perceptual Imaging **5**, 0005021 (2022)
24. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: CVPR. pp. 14318–14328 (2022)
25. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: IPMI. pp. 146–157 (2017)
26. Wei, J., Wang, G., Zhang, S.: Fine-grained medical image out-of-distribution detection through multi-view feature uncertainty and adversarial sample generation. Pattern Recognition p. 111401 (2025)

27. Xia, W., Zhang, Y., Yang, Y., Xue, J.H., Zhou, B., Yang, M.H.: Gan inversion: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(3), 3121–3138 (2022)
28. Xiang, T., Zhang, Y., Lu, Y., Yuille, A.L., Zhang, C., Cai, W., Zhou, Z.: SQUID: Deep feature in-painting for unsupervised anomaly detection. In: CVPR. pp. 23890–23901 (2023)
29. Zavrtanik, V., Kristan, M., Skočaj, D.: DRÆM-a discriminatively trained reconstruction embedding for surface anomaly detection. In: ICCV. pp. 8330–8339 (2021)
30. Zavrtanik, V., Kristan, M., Skočaj, D.: Reconstruction by inpainting for visual anomaly detection. Pattern Recognition **112**, 107706 (2021)
31. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR. pp. 586–595 (2018)
32. Zhang, X., Xu, M., Qiu, D., Yan, R., Lang, N., Zhou, X.: MediCLIP: Adapting CLIP for few-shot medical image anomaly detection. In: MICCAI. pp. 458–468 (2024)
33. Zhang, X., Xu, M., Zhou, X.: RealNet: A feature selection network with realistic synthetic anomaly for anomaly detection. In: CVPR. pp. 16699–16708 (2024)
34. Zhu, J., Yang, C., Shen, Y., Shi, Z., Dai, B., Zhao, D., Chen, Q.: Linkgan: Linking GAN latents to pixels for controllable image synthesis. In: CVPR. pp. 7656–7666 (2023)