City Research Online

City St George's, University of London

Citation: Naval Marimont, S., Siomos, V., Baugh, M., Tzelepis, C., Kainz, B. & Tarroni, G. (2024). Ensembled Cold-Diffusion Restorations for Unsupervised Anomaly Detection. Paper presented at the MICCAI 2024 27th International Conference, 6-10 Oct 2024, Marrakesh, Morocco. doi: 10.1007/978-3-031-72120-5_23

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/34320/

Link to published version: https://doi.org/10.1007/978-3-031-72120-5_23

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to City Research Online policy.

City Research Online: http://openaccess.city.ac.uk/ publications@citystgeorges.ac.uk

arXiv:2407.06635v1 [cs.CV] 9 Jul 2024

Ensembled Cold-Diffusion Restorations for Unsupervised Anomaly Detection

Sergio Naval Marimont[1], Vasilis Siomos[1], Matthew Baugh[2], Christos Tzelepis[1], Bernhard Kainz[2,3], and Giacomo Tarroni[1,2]

[1] City, University of London, UK
[2] Imperial College London, UK
[3] Friedrich–Alexander University Erlangen–Nürnberg, DE
sergio.naval-marimont@city.ac.uk

Abstract. Unsupervised Anomaly Detection (UAD) methods aim to identify anomalies in test samples comparing them with a normative distribution learned from a dataset known to be anomaly-free. Approaches based on generative models offer interpretability by generating anomaly-free versions of test images, but are typically unable to identify subtle anomalies. Alternatively, approaches using feature modelling or self-supervised methods, such as the ones relying on synthetically generated anomalies, do not provide out-of-the-box interpretability. In this work, we present a novel method that combines the strengths of both strategies: a generative *cold-diffusion* pipeline (i.e., a diffusion-like pipeline which uses corruptions not based on noise) that is trained with the objective of turning synthetically-corrupted images back to their normal, original appearance. To support our pipeline we introduce a novel synthetic anomaly generation procedure, called DAG, and a novel anomaly score which ensembles restorations conditioned with different degrees of abnormality. Our method surpasses the prior state-of-the art for unsupervised anomaly detection in three different Brain MRI datasets.

Keywords: Unsupervised anomaly detection · diffusion · synthetic.

1 Introduction

Unsupervised Anomaly Detection (UAD) in medical images involves the detection and/or localization of anomalies without requiring annotations, leveraging only a dataset known to be anomaly-free that *describes* the so-called *normative* distribution. At inference time, UAD methods try to assess whether a given test image belongs to the learnt normative distribution or not. The interest for UAD stems from the fact that differently from common supervised approaches UAD methods 1) do not require annotations, which are expensive and challenging to obtain [14], and 2) are not limited to the labelled classes seen during training.

Multiple strategies have been proposed in the literature. Commonly, UAD methods start by defining a model of the observed normative distribution, either using generative models or self-supervised learning. Methods relying on generative models often *query* the model to produce an image consistent with a given

test image but without anomalies present, and identify as anomalous the areas where the generated image is different from the input test one [3]. While generative models produce a useful healthy counterfactual, their UAD performance is usually limited by blurry reconstructions or lack of coverage of the healthy distribution [11]. Instead, self-supervised strategies typically introduce synthetic anomalies into healthy samples and subsequently train a segmentation network to directly identify the introduced anomalies [21]. Under certain conditions, models are expected to generalize to unseen, naturally occurring anomalies. While very successful in recent editions of MICCAI's Medical Out-of-Distribution Challenge (MOOD)[25], synthetic anomaly models lack in interpretability.

Very recently, a novel strategy called Diffusion-Inspired Synthetic Restoration (DISYRE) [17] combining generative and synthetic anomaly methods have been presented. In this paper, we examine and address the limitations of previously proposed approaches, and implement a hybrid method inspired by diffusion models to turn synthetically corrupted images into healthy ones. Differently from previous approaches, we specifically introduce:

- A novel synthetic anomaly generation process that increases the coverage of the anomaly distribution by disentangling attributes of the generated anomalies, and more importantly, makes the learnt model robust to the assumptions of severity which are required for inference.
- A novel method to identify anomalies which ensembles predictions created under different assumptions on severity.

We evaluated our novel hybrid method named DISYRE v2, in three Brain MRI datasets. DISYRE v2 substantially outperforms all previous methods, setting a new state-of-the-art in UAD. The codebase for DISYRE v2 is available online[4].

2 Related Works

Synthetic anomaly UAD strategies consist in training segmentation networks to identify synthetically generated anomalies, expecting the model to generalize to unseen, naturally occurring abnormalities. First proposed in Foreign Patch Interpolation (FPI) [21], authors realised that the key to generalization is the anomaly generation (AG) procedure, which needs to yield plausible anomalies with a wide coverage of the unknown anomalous distribution. In FPI anomalies are generated by interpolating a squared region of the sample with a patch extracted from a separate sample. A scalar α defines the amount of interpolation used in the corruption and can be used as a proxy for the anomaly severity. A segmentation network is then trained to localize the corruption. Subsequent works increased performance by seamlessly blending the anomalies with the surrounding area [22,25]. More recent works [20,2] improved generalization imposing foreground conditions on the generated anomalies, proposing better training losses for the task and ensembling models trained on different synthetic tasks.

[4] https://github.com/snavalm/disyre.

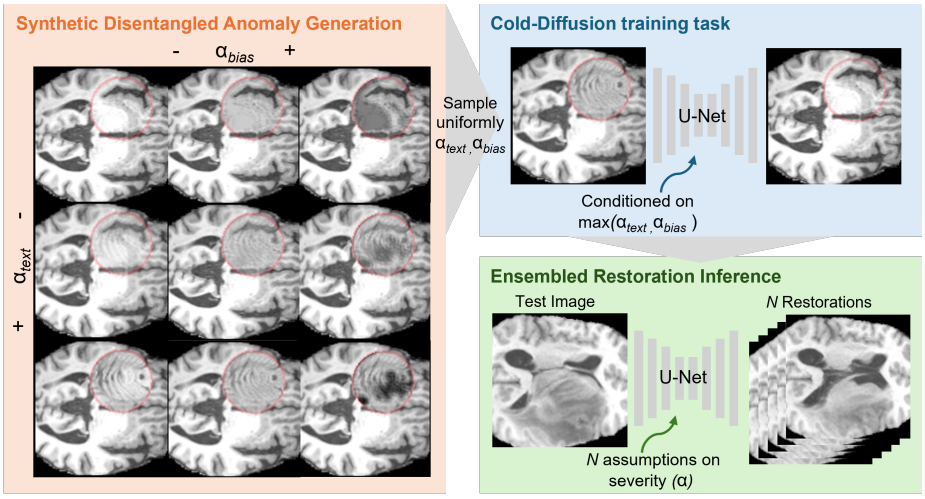

Fig. 1: Overview of the main components of DISYRE v2, our proposed unsupervised anomaly detection pipeline.

On a separate image modelling strategy, diffusion models have recently become the state-of-the-art generative models. Typically, Denoising Diffusion Probabilistic Models (DDPM) [8] are trained to reverse a forward process by which images are gradually corrupted with Gaussian noise. By learning to *undo* the noise corruption, diffusion models learn a score function, $\nabla_x \log p(x)$, which specifies how to modify image pixels to increase their joint probability $p(x)$. Image generation is achieved by iteratively denoising Gaussian noise. Cold-diffusion [1] follows the same concept, but replaces Gaussian noise with alternative corruptions (e.g., blurring, masking). Diffusion models have become recently popular in UAD: the most common approach is to directly train a diffusion model on anomaly-free images. At inference time, test images are corrupted with noise and subsequently denoised, aiming at obtaining an image restoration [4], i.e. an image consistent with the test image but without the original anomalies [24,19]. Anomalies are then identified by comparing test images with their restorations. After corrupting test images with noise, restorations unavoidably shift from the original images, typically producing undesirable changes to low-frequency details, which makes these methods unable to identify subtle anomalies.

Recently an approach named DISYRE [17] showed that, in principle, a score function can act directly as a pixel-wise anomaly score for UAD. Accordingly, a cold-diffusion pipeline with a gradual, synthetic anomaly corruption was proposed, so the learnt score function can generalize to unseen medical anomalies. Similarly to standard diffusion, models are conditioned with a time step t which in DISYRE is a proxy for anomaly severity. Through this conditioning, at inference time DISYRE assumes in practice that test images are highly anomalous, and iteratively uses the modelled score function to generate a restoration. Such

an assumption is not realistic and hampers its ability to identify subtle anomalies, as confirmed by the reported results, where this approach is outperformed by previous methods. Moreover, if the model is instead run assuming lower levels of anomaly severity (acting on the value used for t), the overall performance drops significantly.

3 Methods

3.1 Disentangled Anomaly Generation

We hypothesize that the sensitivity of DISYRE to the assumed severity is due to the choice of FPI [21] as the strategy for generating synthetic anomalies. Since FPI anomalies are created by interpolating two image patches, this is unlikely to produce significant shifts in intensities at the beginning of the corruption process. This feature of FPI can lead to models associating low t only to texture anomalies and, consequently, negatively impacting the restoration abilities of the final model for hyper- and hypo-intense lesions when conditioning with a low t.

To address this issue, we introduce a more expressive Disentangled Anomaly Generation (DAG) process, which disentangles 3 attributes of synthetic anomalies: shape, texture, and intensity bias, where intensity bias is defined as a shift in the expected tissue intensities. When generating anomalies each of the attributes is independently randomly sampled following a uniform distribution and combined to generate a synthetic anomaly. Differently from FPI, DAG can create gradual hyper- and hypo-intense synthetic lesions, which we believe can make the cold-diffusion pipeline robust to the assumed severity during inference. Additionally, by generating anomalies with different combinations of anomalous textures and intensities, DAG can increase the coverage of the anomalous image distribution.

Anomaly Shape: We follow the implementation from [18] where the shape component is determined by a mask $m \in \mathbb{R}^{H \times W}$, with H and W respectively height and width of the image, and $m_i \in [0, 1]$, where pixels with $m_i = 0$ are uncorrupted and $m_i = 1$ are anomalous. Masks are generated randomly and their edges smoothed so anomalies blend gradually with surrounding areas.

Anomaly Texture: The texture component is created following FPI [21]. Image patches randomly sampled from the healthy training set, referred to as *foreign patches* $x_{fp} \in \mathbb{R}^{H \times W}$, are used to replace sections of healthy training images $x \in \mathbb{R}^{H \times W}$. The corruption is achieved with a convex combination of x and x_{fp} governed by the interpolation factor $\alpha_{text} \in [0, 1]$. Differently from FPI, we attempt to separate bias and texture components by normalizing x_{fp} so it has the same range of intensities as the original image x in the anomaly location (being $\overline{x}_{fp}$ normalized version of x_{fp}). Texture corrupted images x_{text} are defined as:

$$x_{text} = (1 - \alpha_{text} \cdot m) \cdot x + \alpha_{text} \cdot m \cdot \overline{x}_{fp} \tag{1}$$

Intensity Bias: Intensity corruptions are achieved shifting intensities in the location determined by m by a factor $\pm\alpha_{bias}$. To obtain more plausible anomalies

we aim to restrict the shift of intensities to specific tissue types (e.g. white or gray matter in the case of Brain MRI) within the anomaly location determined by m. We propose to use k-Means to identify clusters of intensities across the whole dataset which we assume to represent tissue types ($K = 5$ clusters are used in our experiments). When generating anomalies, we randomly choose the tissue type k and define a tissue type mask as $m_{tissue} \in \mathbb{R}^{H \times W}$ with elements $m^i_{tissue} = 1$ if x^i is assigned to cluster k and $m^i_{tissue} = 0$ otherwise. Additionally, the randomly sampled $bias_{sign} \in \{-1, 1\}$ defines whether intensities will be decreased or increased. The bias corruption is applied to x_{text} to obtain the final synthetically corrupted image x_{sc} used during training:

$$x_{sc} = (1 - m \cdot m_{tissue}) \cdot x_{text} + (x_{text} + \alpha_{bias} \cdot bias_{sign}) \cdot m \cdot m_{tissue} \quad (2)$$

Figure 1 shows anomalies generated when varying attributes α_{bias} and α_{text}.

3.2 Cold-diffusion with synthetic anomaly corruptions

In diffusion models [8], a corruption schedule $B(t)$ defines a gradual Gaussian noise corruption process which starts at time step $t = 0$, where images are unaltered, and ends at $t = T$, where only Gaussian noise remains. Diffusion models are trained to reverse this process: by denoising samples while conditioned on t, they learn to generate high-frequency details for low values of t and structural features for high values. Cold-diffusion [1] follows the same strategy but proposes alternative corruptions not based on noise.

We implement a synthetic anomaly cold-diffusion method to generate restorations from images with anomalies following DISYRE [17]. In the original implementation, the noise schedule $B(t)$ from DDPMs is adopted and used to set the FPI [21] interpolation factor $\alpha = B(t)$. Consequently, when $t = 0$, $\alpha_t = 0$ the image belongs to the observed healthy distribution, while when $t = T$, $\alpha_t = 1$ and the image is highly anomalous. A network P_θ is trained to restore corrupted images x_t into their original counterparts x_0 when conditioning on t, $x_0 \approx \hat{x}_0 = P_\theta(x_t, t)$. Network parameters are therefore optimized using the objective:

$$\mathbb{E}_{t \sim [1,T], x_0, x_{fp}}(\|x_0 - P_\theta(x_t, t)\|^2) \quad (3)$$

We propose to extend the model to the 2 dimensions of corruption in DAG, i.e. α_{text} and α_{bias}, by simply conditioning on $t = B^{-1}\left(max(\alpha_{text}, \alpha_{bias})\right)$, where B^{-1} is the inverse of $B(t)$, so the model receives a single indication of severity.

3.3 Anomaly localization

Diffusion models generate images iteratively, starting at step $t = T$ with Gaussian noise $x_T \sim \mathcal{N}(0, I)$, where $x_T \in \mathbb{R}^{H \times W}$. At each of the T iterations the diffusion model *denoises* the sample until obtaining an image from the observed distribution at step $t = 0$. Similarly, [17] proposes to iteratively use the score function learned in the domain of synthetically corrupted images to *heal* test

images x and obtain restorations. The residuals between restorations and original images are then used as anomaly score (AS). Importantly, by starting the restoration process with $t = T$ the method assumes a high severity of abnormality. We refer to the AS proposed in [17] as *multi-step* in our experiments.

In this work, we propose an effective alternative, which is to ensemble single-step restorations conditioned at different degrees of abnormality.

$$AS_{ensemble} = \sum_{t=T}^{0} (\|P_\theta(x, t) - x\|) \tag{4}$$

Our proposed Ensemble AS allows us to leverage the increased model robustness from DAG. Similarly to [17], in practice we skip steps based on a *step size* hyper-parameter. We choose $step_size = 25$ as a compromise between coverage of the anomalous distribution and inference speed, and report results in the supplementary materials. As in [17], we use sliding window inference and weight the AS predicted for individual patches using the proportion of foreground. Sliding window inference allows to apply our pipeline seamlessly to other high-dimensional modalities. In our experiments we use a patch-overlap of 0.25 and *gaussian weighting* of patches [9].

4 Experiments

4.1 Experimental setup

Our experiments use the Brain MRI setup from [11] which includes datasets:

- Cambridge Centre for Ageing and Neuroscience (CamCAN) [23]: 653 T1- and T2-weighted images of healthy and lesion free subjects. CamCAN dataset is used for training ($N = 603$) and to evaluate training convergence ($N = 50$).
- Anatomical Tracings of Lesions After Stroke (ATLAS) dataset [13]: 655 T1-weighted images of stroke patients with annotated lesions. Some of these images are affected by artefacts and inadequate skull-stripping, which UAD methods often identify as anomalies despite not being annotated as such and thus lead to performance underestimation. For our internal ablations we use the subset excluding issues (ATLAS-ex, N=571), but report the results for the full dataset when comparing to previous studies [11].
- Multimodal Brain Tumor Segmentation (BraTS) Challenge dataset [16] 2020 edition: 369 T1- and T2-weighted images from subjects with gliomas, provided with manual annotations of lesions.

ATLAS and BraTS datasets are used to evaluate UAD performance while CamCAN is used as the training, normative dataset. We follow the pre-processing specified in [11]. Similarly to [17], we used a 2D patch-based pipeline with patches of 128×128 size obtained from non-empty axial slices. Our network implementation, diffusion corruption schedule, and training schedule follow the specifications in [17] but we ran training for longer (200,000 steps) to ensure convergence.

Table 1: Localization results of the image-reconstruction (**IR**), feature-modeling (**FM**), self-supervised (**S-S**) and hybrid (**HY**) methods. Best scores are bold.

		ATLAS		**BraTS-T1**		**BraTS-T2**	
	Method	AP	⌈Dice⌉	AP	⌈Dice⌉	AP	⌈Dice⌉
IR	VAE [3]	0.11	0.20	0.13	0.19	0.28	0.33
	r-VAE [4]	0.09	0.17	0.13	0.19	0.36	0.40
FM	FAE [15]	0.08	0.18	0.42	0.45	0.51	0.52
	PaDiM [5]	0.05	0.13	0.21	0.28	0.34	0.38
	CFLOW-AD [7]	0.04	0.10	0.16	0.24	0.31	0.35
	RD [6]	0.11	0.22	0.36	0.42	0.47	0.50
S-S	PII [22]	0.03	0.07	0.13	0.22	0.13	0.22
	DAE [10]	0.05	0.13	0.13	0.20	0.47	0.49
	CutPaste [12]	0.03	0.06	0.07	0.13	0.22	0.26
	MOOD22 [18]	0.10	0.21	0.24	0.31	0.47	0.48
HY	DISYRE [17]	0.29	0.37	0.34	0.41	0.75	0.70
	DISYRE v2 (Ours)	**0.33**	**0.45**	**0.48**	**0.51**	**0.79**	**0.73**

4.2 Results and Discussion

To quantitatively evaluate our method, we use Average Precision (AP) and an estimate of the best possible Sørensen-Dice index (⌈Dice⌉) as in [11]. To evaluate the robustness of the model to the assumed severity (t), we obtained AS from single-step restorations, i.e. $\|x - P(x,t)\|$ for different t. Figure 2 shows that DAG improves results across all values of t. The improvement is particularly relevant when conditioning with $t = 19$ and $t = 39$ (i.e. low severity assumed) in BraTS-T2 and ATLAS datasets. Since anomalies in BraTS-T2 and ATLAS appear hyper- and hypo-intense respectively, these results showcase that, by enabling the creation of low-grade bias-only anomalies, DAG addresses the limitations of FPI and makes the model robust across different severity levels.

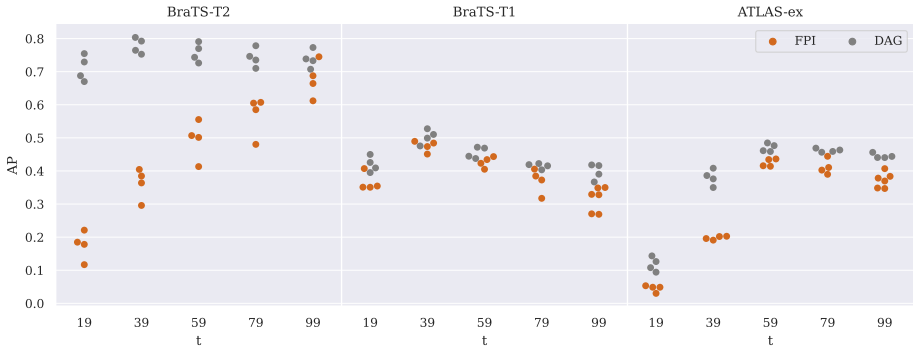

Fig. 2: AP for 4 independent seeds, using single-step restoration, FPI vs DAG.

Table 2: AP in test sets for combinations of Anomaly Score (AS) and Anomaly Generation (AG). Mean and std. dev. on 4 seeds.

AS	AG	**ATLAS-ex**	**BraTS-T1**	**BraTS-T2**
Unconditional	FPI	0.36 ± 0.03	0.44 ± 0.01	0.68 ± 0.00
Unconditional	DAG	0.40 ± 0.01	0.44 ± 0.01	0.74 ± 0.01
Multi-step	FPI	0.42 ± 0.02	0.40 ± 0.02	0.73 ± 0.04
Multi-step	DAG	0.47 ± 0.00	0.44 ± 0.02	0.78 ± 0.02
Ensemble	FPI	0.44 ± 0.02	0.46 ± 0.02	0.69 ± 0.06
Ensemble	DAG	**0.47± 0.01**	**0.48 ± 0.01**	**0.79 ± 0.02**

Next we evaluated the interaction between proposed AG and AS strategies in Table 2. To further evaluate the impact of conditioning, we also include models trained to generate restorations without conditioning on t altogether, with their AS defined as $\|x - P(x)\|$. Conditional setups show overall better performance compared to unconditional ones, showcasing the benefits of the diffusion-like pipeline. However, prior methods (i.e. multi-step AS) fail to improve the unconditional setting in BraTS-T1 (where anomalies are often subtle) by assuming a high severity during inference time. DISYRE v2 solves this limitation with the proposed Ensemble AS, which raises the AP by 10% (0.48 vs 0.44). The interaction between the proposed AS-AG strategies is noteworthy: DAG makes single-step restorations robust to the assumed severity, benefiting the Ensembled AS when combining predictions.

Finally, Table 1 includes a comparison of our method with baselines evaluated in [11] and DISYRE [17]. The performance of hybrid methods showcases that the combination of image restoration and anomaly localization tasks improve above their individual components, *i.e.* image-reconstruction and synthetic-anomaly localization [22,12]. By addressing the described limitations of the original DISYRE, DISYRE v2 outperforms it by a good margin across the 3 tasks. Figure 3 includes qualitative examples of predictions. Additional comparisons between Ensemble and Multi-step AS can be found in the supplementary materials.

5 Conclusion

In this work we presented DISYRE v2, a novel evolution of DISYRE, and evaluated its performance in a common test bench of Brain MRI lesions. We set the new state-of-the-art in the three tasks where we evaluated it, and importantly, improved the performance metrics for the two more challenging tasks (ATLAS and BraTS-T1) by more than 10%. In our analysis we identify limitations of previous works and present as a solution a novel combination of synthetic AG, AS and cold-diffusion modelling which allows to make fewer assumptions at inference time, removing biases towards specific anomalies. Consequently, DISYRE v2 not only improves performance in the tasks evaluated, but it is expected to be more robust across the diverse and challenging naturally occurring medical anomalies, setting a new state-of-the-art in medical UAD.

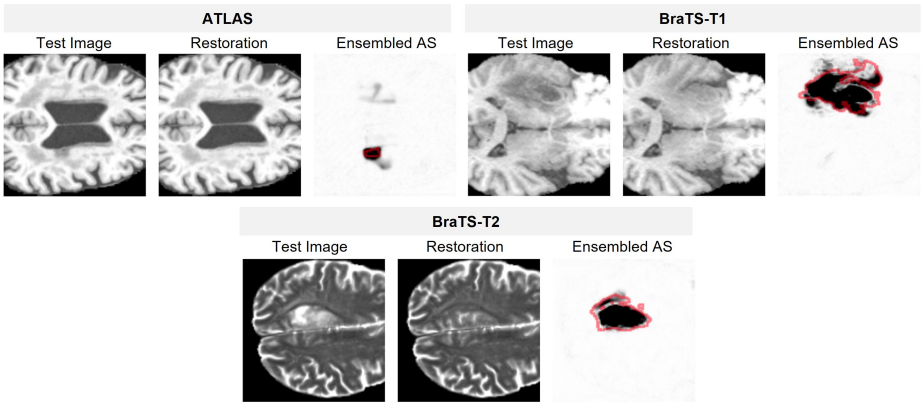

Fig. 3: Examples of restorations and Ensembled AS vs ground truth (red outline).

Acknowledgments: B. Kainz received funding from the ERC project MIA-NORMAL 101083647; M. Baugh is supported by a UKRI DTP award.

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

References

1. Bansal, A., Borgnia, E., Chu, H.M., Li, J.S., Kazemi, H., Huang, F., Goldblum, M., Geiping, J., Goldstein, T.: Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise (Aug 2022), arXiv:2208.09392 [cs]
2. Baugh, M., Tan, J., Müller, J.P., Dombrowski, M., Batten, J., Kainz, B.: Many Tasks Make Light Work: Learning to Localise Medical Anomalies from Multiple Synthetic Tasks. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 162–172. Lecture Notes in Computer Science, Cham (2023)
3. Baur, C., Denner, S., Wiestler, B., Albarqouni, S., Navab, N.: Autoencoders for Unsupervised Anomaly Segmentation in Brain MR Images: A Comparative Study. Medical image analysis (02 Jan 2021), 69:101952 (2021)
4. Chen, X., You, S., Tezcan, K.C., Konukoglu, E.: Unsupervised lesion detection via image restoration with a normative prior. Medical Image Analysis **64**, 101713 (Aug 2020)
5. Defard, T., Setkov, A., Loesch, A., Audigier, R.: PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. In: Del Bimbo, A., Cucchiara, R., Sclaroff, S., Farinella, G.M., Mei, T., Bertini, M., Escalante, H.J., Vezzani, R. (eds.) Pattern Recognition. ICPR International Workshops and Challenges. pp. 475–489. Lecture Notes in Computer Science, Cham (2021)
6. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9737–9746 (2022)

7. Gudovskiy, D., Ishizaka, S., Kozuka, K.: CFLOW-AD: Real-Time Unsupervised Anomaly Detection with Localization via Conditional Normalizing Flows. In: 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1819–1828. IEEE Computer Society, Los Alamitos, CA, USA (Jan 2022)
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
9. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021), publisher: Nature Publishing Group
10. Kascenas, A., Pugeault, N., O'Neil, A.Q.: Denoising autoencoders for unsupervised anomaly detection in brain mri. In: International Conference on Medical Imaging with Deep Learning. pp. 653–664. PMLR (2022)
11. Lagogiannis, I., Meissen, F., Kaissis, G., Rueckert, D.: Unsupervised Pathology Detection: A Deep Dive Into the State of the Art. IEEE Transactions on Medical Imaging pp. 1–1 (2023), arXiv:2303.00609 [cs]
12. Li, C.L., Sohn, K., Yoon, J., Pfister, T.: Cutpaste: Self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9664–9674 (2021)
13. Liew, S.L., Anglin, J.M., Banks, N.W., Sondag, M., Ito, K.L., Kim, H., Chan, J., Ito, J., Jung, C., Khoshab, N., others: A large, open source dataset of stroke anatomical brain images and manual lesion segmentations. Scientific data **5**, 1–11 (2018)
14. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I.: A Survey on Deep Learning in Medical Image Analysis. Medical image analysis **vol. 42**, 60–88 (2017)
15. Meissen, F., Paetzold, J., Kaissis, G., Rueckert, D.: Unsupervised Anomaly Localization with Structural Feature-Autoencoders. In: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 14–24. Cham (2023)
16. 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)
17. Naval Marimont, S., Baugh, M., Siomos, V., Tzelepis, C., Kainz, B., Tarroni, G.: DISYRE: Diffusion-inspired synthetic restoration for unsupervised anomaly detection. In: Proceedings/IEEE International Symposium on Biomedical Imaging (ISBI). IEEE (2024)
18. Naval Marimont, S., Tarroni, G.: Achieving state-of-the-art performance in the Medical Out-of-Distribution (MOOD) challenge using plausible synthetic anomalies (Nov 2023), arXiv:2308.01412 [cs]
19. Pinaya, W.H.L., Graham, M.S., Gray, R., da Costa, P.F., Tudosiu, P.D., Wright, P., Mah, Y.H., MacKinnon, A.D., Teo, J.T., Jager, R., Werring, D., Rees, G., Nachev, P., Ourselin, S., Cardoso, M.J.: Fast Unsupervised Brain Anomaly Detection and Segmentation with Diffusion Models. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 705–714. Lecture Notes in Computer Science, Cham (2022)
20. Schlüter, H.M., Tan, J., Hou, B., Kainz, B.: Natural Synthetic Anomalies for Self-supervised Anomaly Detection and Localization. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) Computer Vision – ECCV 2022. pp. 474–489. Lecture Notes in Computer Science, Cham (2022)
21. Tan, J., Hou, B., Batten, J., Qiu, H., Kainz, B., et al.: Detecting outliers with foreign patch interpolation. MELBA **1**(April 2022 issue), 1–27 (2022)

22. Tan, J., Hou, B., Day, T., Simpson, J., Rueckert, D., Kainz, B.: Detecting Outliers with Poisson Image Interpolation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Cham (2021)
23. Taylor, J.R., Williams, N., Cusack, R., Auer, T., Shafto, M.A., Dixon, M., Tyler, L.K., Henson, R.N., others: 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)
24. Wyatt, J., Leach, A., Schmon, S.M., Willcocks, C.G.: AnoDDPM: Anomaly Detection with Denoising Diffusion Probabilistic Models using Simplex Noise. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 649–655. New Orleans, LA, USA (Jun 2022)
25. Zimmerer, D., Full, P.M., Isensee, F., Jager, P., Adler, T., Petersen, J., Kohler, G., Ross, T., Reinke, A., Kascenas, A., Jensen, B.S., O'Neil, A.Q., Tan, J., Hou, B., Batten, J., Qiu, H., Kainz, B., Shvetsova, N., Fedulova, I., Dylov, D.V., Yu, B., Zhai, J., Hu, J., Si, R., Zhou, S., Wang, S., Li, X., Chen, X., Zhao, Y., Marimont, S.N., Tarroni, G., Saase, V., Maier-Hein, L., Maier-Hein, K.: MOOD 2020: A Public Benchmark for Out-of-Distribution Detection and Localization on Medical Images. IEEE transactions on medical imaging **41**(10), 2728–2738 (Oct 2022)