

Panda: Unsupervised Pelvic Anomaly Detection for Real-Time MR Imaging

Anika Knupfer^{1,2}, Maximilian Lindholz^{3,4}, Johanna Paula Müller⁵, Jordina Aviles Verdera^{1,2}, Smiti Tripathy⁶, Susanne Schulz-Heise⁶, and Jana Hutter^{1,2}

¹ Institute of Information Processing, Leibniz University Hannover, Hannover, Germany

² CAIMED, L3S, Hannover, Germany

³ Department of Radiology, Charité Universitätsmedizin Berlin, Berlin, Germany

⁴ Berlin Institute of Health, Berlin, Germany

⁵ Image Data Exploration and Analysis Lab, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany

⁶ Institute for Radiology, University Hospital Erlangen, Erlangen, Germany
`knupfer@tnt.uni-hannover.de`

Abstract. Female pelvic diseases remain an under researched area characterized by often delayed diagnosis. While pelvic MRI offers superior soft-tissue contrast for diagnosis and image-guided procedures, real-time anomaly detection remains challenging due to physiological motion, tissue deformation, and instrument artifacts. Existing supervised approaches are impractical, as adverse events are rare, heterogeneous, and difficult to annotate. We present a Dinomaly-based unsupervised anomaly detection framework adapted for pelvic MRI that learns normative representations from healthy cases and flags deviations without requiring labels. Our approach leverages a frozen DINOv3 Vision Transformer encoder combined with a noisy MLP bottleneck and Linear Attention decoder to prevent identity mapping while maintaining computational efficiency. Anomalies are localized via per-token cosine distance between encoder and decoder representations, yielding spatial anomaly maps that provide immediate feedback at the scanner to support radiologist decision-making and adaptive protocol adjustment. Evaluated on a curated subset of the Uterine Myoma Dataset, the framework achieves a pixel-level AUROC of 88.06% and high specificity (95.45%) at frame level at 40.5 slices/s, meeting real-time clinical deployment requirements. The spatial anomaly maps and frame-level scores provide immediate, localized feedback at the scanner to support radiologist decision-making and adaptive protocol adjustment during active procedures.

Keywords: Unsupervised Anomaly Detection · Abdominal MRI · Vision Transformers · Pelvic Imaging · Foundation Models

1 Introduction

Female pelvic diseases such as adenomyosis, endometriosis, cervical and endometrial cancer affect a significant number of women but remain underdiagnosed,

with for example endometriosis affecting 10% of women of reproductive age but suffering from an average time-till-diagnosis of 8 years. The complex anatomy, physiological motion, high variability and proximity of multiple organs are significant challenges both for diagnostic imaging and for increasingly available image-guided interventions [12,17,26,28]. MRI plays thereby a growing role [6,23] and is ideally suited particularly to guide interventions such as biopsies, ablations and brachytherapy in the anatomically complex and motion-prone pelvic region [19,25]. However, despite these benefits carrying the promise of accelerating diagnosis and minimizing treatments side-effects, pelvic MRI remains challenging due to substantial physiological motion from respiration, peristalsis, and bladder filling [3] and the lack of interactivity in conventional MRI examinations [6,10]. The general paucity of well-curated data and labels in this under-researched area exhibiting substantial physiological variability further aggravates the problem [1,4].

Classical supervised machine learning approaches are particularly ill-suited, as many clinically relevant anomalies are rare, heterogeneous, and difficult to label [18]. Moreover, for interventional settings, it is impractical to collect large, well-annotated datasets representing adverse events [21]. This motivates the use of unsupervised anomaly detection, where models learn a representation of “normal” pelvic anatomy and dynamics from routine cases and flag deviations without requiring explicit labels [8]. Recent advances in deep learning, spatio-temporal modeling, and generative models can potentially detect subtle deviations that may be invisible to the naked eye or that emerge gradually over time [9,16]. However, translating these techniques to pelvic MRI poses unique challenges and requires due to the inherent data paucity bespoke methods [7,21].

This work contributes (1) a novel framework for unsupervised pelvic anomaly detection, systematically redesigning the Dinomaly architecture [2] for this domain, directly addressing the annotation bottleneck that renders supervised approaches impractical in this setting, (2) a reproducible evaluation benchmark for unsupervised pelvic MRI anomaly detection, and provide the first head-to-head comparison between reconstruction-based and generative detection paradigms, revealing complementary strengths that inform future system design and (3) clinically actionable real-time performance, delivering simultaneous frame-level slice flagging and spatially precise pixel-level anomaly maps immediately available at the scanner to adapt acquisition protocols immediately.

2 Theory

2.1 Problem Formulation

We formulate anomaly detection as an unsupervised learning problem. $\mathcal{D}_{\text{train}} = \{x_i\}_{i=1}^N$ are the set of N training frames extracted from procedurally normal pelvic MRI sequences, acquired in the absence of pathologies, major artifacts,

adverse events and clinically significant deviations. The model learns a representation of normal pelvic appearance, such that deviating inputs induce elevated reconstruction error. Anomaly localization is performed via per-pixel cosine dissimilarity between encoder and decoder representations. Each frame $x \in \mathbb{R}^{H \times W}$ is a 2D sagittal slice, and the model produces a pixel-level anomaly map $\mathcal{A}(x) \in [0, 1]^{H \times W}$, where $\mathcal{A}(x)_{h,w}$ quantifies local deviation from normality at position (h, w) for $h \in \{1, \dots, H\}$, $w \in \{1, \dots, W\}$. The frame-level anomaly score $s(x) = \max_{h,w} \mathcal{A}(x)_{h,w} \in [0, 1]$ captures the most pronounced anomalous responses within the frame, providing a global slice classification. Inference operates under near real-time constraints.

2.2 Dinomaly-Based Architecture

Our approach extends Dinomaly [2], a minimalist reconstruction-based anomaly detection architecture built exclusively from Transformer blocks [24]. Operating at the representation level improves robustness to the pronounced intensity variability typical for MR imaging. The encoder is a pre-trained foundation model, a Vision Transformer (ViT). It remains entirely frozen during training and inference, with feature maps from intermediate layers extracted as reconstruction targets. Freezing prevents encoder-decoder co-adaptation and preserves the pre-trained representation space as a stable normative reference. The bottleneck is a MLP with training-time Dropout that introduces stochastic feature corruption, limiting the decoder’s generalization to unseen inputs to avoid the reconstruction of anomalous patterns it was never trained on [27]. By corrupting aggregated encoder features during training, the bottleneck forces the decoder to restore normal structure from incomplete information, ensuring elevated reconstruction error on anomalous inputs at inference. Dropout is disabled at inference for deterministic operation. The decoder replaces standard Softmax attention with Linear Attention (LA) [5] to promote globally distributed feature aggregation over spatially concentrated token interactions, well-suited to pelvic MRI where anomalies manifest as subtle contextual inconsistencies. LA is defined as

$$\text{LA}(Q, K, V) = \phi(Q)(\phi(K)^\top V), \quad \phi(\cdot) = \text{elu}(\cdot) + 1, \quad (1)$$

where Q, K, V are the query, key, and value matrices, and $\phi(\cdot)$ is an element-wise non-negative activation enabling the kernel factorization, favoring globally consistent reconstruction over strict token-wise correspondence. Training minimizes a group-to-group cosine reconstruction loss. Encoder feature maps are partitioned into two groups capturing complementary levels of abstraction: low-level features (layers 3-6), encoding local texture and high-level features (layers 7-10), encoding semantic context. The decoder is trained to reconstruct each group independently, with gradient shrinking to emphasize hard tokens and sharpen sensitivity to subtle deviations. Denoting the flattened grouped encoder and decoder feature maps as $F(f_E)$ and $F(\hat{f}_D)$, the loss is

$$\mathcal{L} = D_{\cos}(F(f_E), F(\hat{f}_D)), \quad D_{\cos}(a, b) = 1 - \frac{a^\top b}{\|a\| \|b\|}, \quad (2)$$

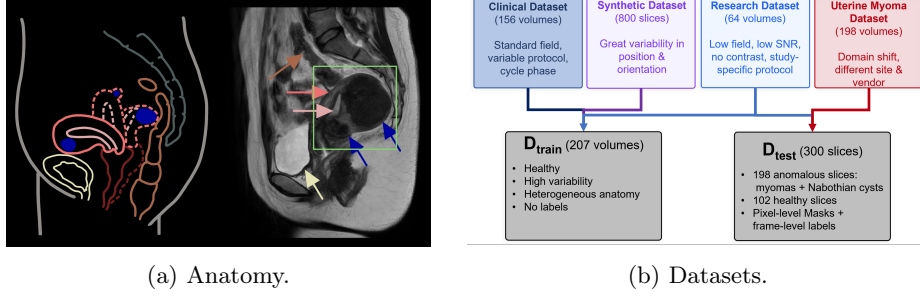


Fig. 1: (a) Schematic and T2w MRI pelvic illustration (pink—uterus; light pink—endometrium; blue—myoma; yellow—bladder; brown—rectum. Dashed lines indicate physiological uterine variability and the green bounding box the uterine crop. (b) Data flowchart illustrating the assembly of $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{test}}$.

3 Method

3.1 Datasets

Training Data. $\mathcal{D}_{\text{train}}$ comprises 207 anonymized sagittal T2w healthy MRI volumes (see Fig. 1a, ethics numbers: 24-304-Br, 23-444-Bm), assembled from a clinical and a research dataset supplemented with synthetic data (Fig. 1b). The clinical dataset with 156 volumes, employed with institutional ethics approval exhibits substantial field strength, protocol, usage of contrast and physiological variability. The research dataset, consisting of 51 volumes, was acquired at low field (0.55T) and is hence characterized by low SNR. The 800 publicly available synthetic images [13] covering especially rare anatomical variants were reviewed by radiologists for anatomical plausibility and verified against existing training samples using cosine similarity (threshold 0.95) to avoid near-duplicate images.

Evaluation Data and out-of-distribution evaluation. $\mathcal{D}_{\text{test}}$ combines 198 anomalous volumes from the public Uterine Myoma Dataset (UMD) [14, 15] and 13 low field research dataset volumes held out for frame-level evaluation. The UMD dataset was acquired at a different site with a different vendor and greatly varying parameters and thus constitutes a genuine out-of-distribution evaluation.

Data Preparation. All volumes were resampled to $0.5 \times 0.5 \times 1 \text{ mm}^3$, resized to $256 \times 256 \times 30$, and intensity-normalized to $[0, 1]$. For $\mathcal{D}_{\text{train}}$, masking was performed with an attention-based 3D U-Net [22], for $\mathcal{D}_{\text{test}}$ provided masks were used. Bounding boxes of size $96 \times 96 \times n$ ($n = \text{slices containing uterine tissue}$) were extracted (See Fig. 1a). Synthetic training images required no further spatial preprocessing. For evaluation, one slice per examination containing both annotated anomalous regions and visible uterine structures was selected, yielding 198 anomalous slices. The 102 anomaly-free slices were sampled at a minimum distance of 2 slices from any annotated lesion (UMD) or with a minimum inter-slice spacing of 2 (held-out volumes) to avoid similar neighboring frames.

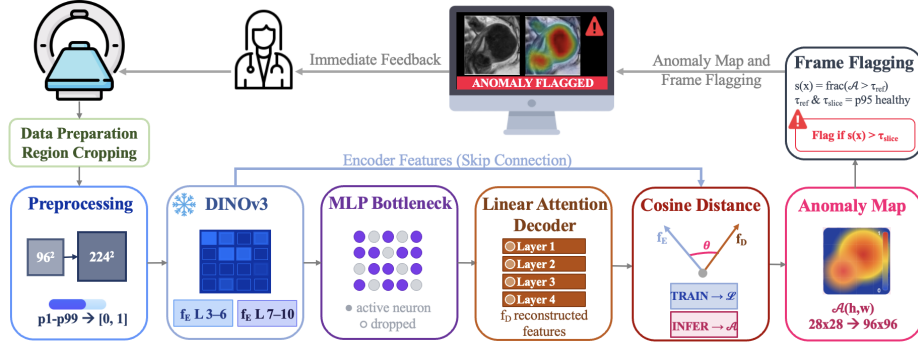


Fig. 2: Proposed framework for low-latency pelvic anomaly detection in MRI.

3.2 Architecture and Training.

The encoder is a frozen DINOv3 ViT-L/16 [11, 20], selected to balance representational capacity and overfitting risk in a low-data setting. Input slices are resized from 96×96 to 224×224 and normalized using per-slice percentile scaling to $[0, 1]$ (1st-99th percentile) followed by standard ImageNet normalization, producing patch-level embeddings of dimension $d = 1024$ on a 28×28 grid. The bottleneck MLP applies dropout ($p = 0.136$) during training, and the decoder consists of four Linear Attention layers with eight attention heads and an MLP ratio of 2.097. The anomaly map $\mathcal{A}(h, w) = D_{\cos}(f_E, \hat{f}_D)$ is computed via a skip connection bypassing the bottleneck and upsampled from 28×28 to 96×96 for evaluation. The model is trained for 120 epochs with batch size 32, learning rate 4.24×10^{-4} , and weight decay 10^{-4} using AdamW, without data augmentation. Hard-token weighting via gradient shrinking is applied from the first epoch at the 33rd percentile of per-token reconstruction error, down-weighting well-reconstructed tokens by a factor of 0.1. Hyperparameters were selected via Bayesian optimization over 40 configurations directly maximizing pixel-level AUROC on $\mathcal{D}_{\text{test}}$, with no gradient updates derived from test data, as validation loss showed negligible correlation with detection performance ($\rho_S = +0.094$, $p = 0.566$). The full pipeline is illustrated in Fig. 2 including the in-built feedback mechanism available for instant feedback at the time of acquisition.

3.3 Evaluation.

Pixel-Level Anomaly Scoring is performed on the continuous anomaly maps $\mathcal{A}(h, w)$ using AUROC. For binary metrics, a pixel threshold τ_{px} is determined by maximizing the Dice coefficient across $\mathcal{D}_{\text{test}}$, providing an upper bound on performance. **Frame-Level Anomaly Scoring** reduces the anomaly map $\mathcal{A}$ to a scalar score measuring the spatial extent of anomalous responses to capture the prevalence of atypical responses, avoiding the noise sensitivity of maximum-

based scores and the dilution of localized anomalies in mean-based scores:

$$s(x) = \frac{1}{HW} \sum_{h,w} \mathbf{1}[\mathcal{A}(h, w) > \tau_{\text{ref}}]. \quad (3)$$

Both τ_{ref} and the slice-level decision threshold τ_{slice} are calibrated exclusively on healthy slices from $\mathcal{D}_{\text{train}}$. τ_{ref} is set to the 95th percentile of pixel-level anomaly scores and τ_{slice} to the 95th percentile of $s(x)$. A frame is flagged as anomalous at inference if $s(x) > \tau_{\text{slice}}$.

4 Experiments and Results

We evaluate the proposed method on $\mathcal{D}_{\text{test}}$ at pixel-level and frame-level, and benchmark against a current state of the art ResVAE [8].

Anomaly Detection Performance. Qualitatively, 4 representative examples (Fig. 3) display anomaly maps corresponding closely with GT. Two failure patterns emerge with smaller or low-contrast lesions (case 4) and adjacent structures like the bowel occasionally causing false positives (case 2). Results comparing upscaled 96×96 anomaly maps with expert-annotated segmentation masks are listed in Table 1. The pixel-level AUROC of 88.06% demonstrates strong detection ability. At the retrospectively determined optimal threshold, accuracy reaches 93.62% and specificity 95.10%, demonstrating that the vast majority of normal pixels are classified correctly. The low sensitivity of 38.49% and precision of 17.31% is linked to the strong class imbalance: pathological areas occupying only a small fraction of pixels results in even spatially well-localized anomaly maps demonstrating relatively low precision and sensitivity on a pixel-by-pixel basis. The F1 score of 23.87% should be interpreted in this context. Taken together, the high AUROC value and high specificity suggest that the model generates spatially meaningful anomaly maps, while the precise delineation of lesion boundaries remains challenging. In frame-level results, the healthy-calibrated threshold τ_{slice} yields high specificity (95.45%) and precision (71.88%) (Table 1). The low sensitivity (11.61%) is a direct consequence of the conservative threshold and the coarse token resolution where small lesions may not elevate a sufficient fraction of pixels to trigger a flag. The frame-level AUROC of 68.41% confirms moderate discriminative ability, and the high-specificity operating point reflects a design choice prioritizing trust over exhaustive recall. Per-slice inference latency, measured on the UMD test set using an NVIDIA GeForce RTX 3080 (10 GB), was 0.0247 s per slice or 40.5 slices/s, consistent with real-time requirements for MRI workflows.

Comparison with ResVAE Baseline. The re-trained ResVAE [8] achieves markedly lower pixel-level AUROC (69.65 vs. 87.85) and precision (6.95 vs. 17.04), but substantially higher sensitivity (79.91 vs. 38.39), reflecting its tendency to produce diffuse, high-recall anomaly maps at the cost of poor specificity (64.43 vs. 95.02). Figure 4 illustrates the complementary nature of the two approaches across two

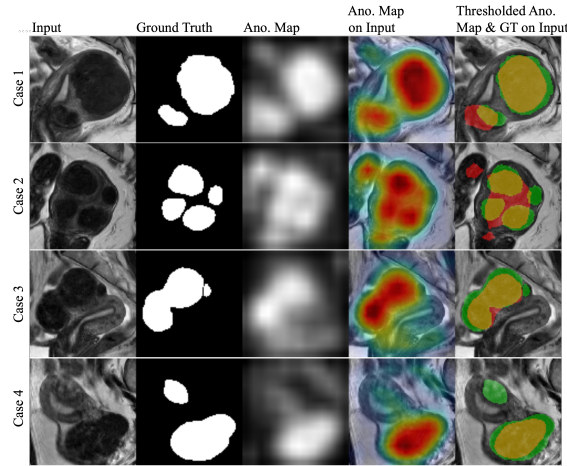


Fig. 3: Representative results on $\mathcal{D}_{\text{test}}$. Each row presents from left to right: T2w input slice, GT, raw anomaly map, heatmap overlay on the input, and threshold-based prediction maximizing the Dice across $\mathcal{D}_{\text{test}}$ overlaid with GT on the input (yellow: predicted anomaly area, green: GT boundary, red: false positive).

Table 1: Detection performance (%) on $\mathcal{D}_{\text{test}}$, comparing the proposed Dinomaly-based architecture with the state-of-the-art ResVAE [8]. Best scores in bold.

Architecture	Evaluation	AUROC $\uparrow$	Accuracy $\uparrow$	Precision $\uparrow$	Sensitivity $\uparrow$	Specificity $\uparrow$	F1 $\uparrow$
Dinomaly-based	Frame-level	68.41	53.53	71.88	11.61	95.45	-
	Pixel-level	88.06	93.62	17.31	38.49	95.10	23.87
SotA ResVAE [8]	Pixel-level	69.65	64.63	6.95	79.91	64.43	11.21

representative cases: the reconstruction-based ResVAE method responds to subtle low-contrast lesions with greater sensitivity (white arrows) while Dinomaly provides spatially precise localization of dominant anomalous structures, consistent with the semantic representations of the frozen DINOv3 encoder. The coarser 28×28 token grid remains the primary limitation of the proposed method with respect to small lesion sensitivity.

Ablation Study. The small differences between models (Table 2) reflect the robustness of the frozen DINOv3 encoder, whose representations are largely insensitive to reconstruction decisions. Individual modifications involve trade-offs (e.g., removing dropouts improves accuracy/specificity, while ViT-S/16 and larger decoders increase sensitivity but carry the risk of overfitting). The full configuration was selected based on the best overall balance. Replacing ViT-L/16 with ViT-S/16 reduces the AUROC by 4.1% but significantly lowers inference latency (76.8 slices/s). Training with 50% results in significant deterioration.

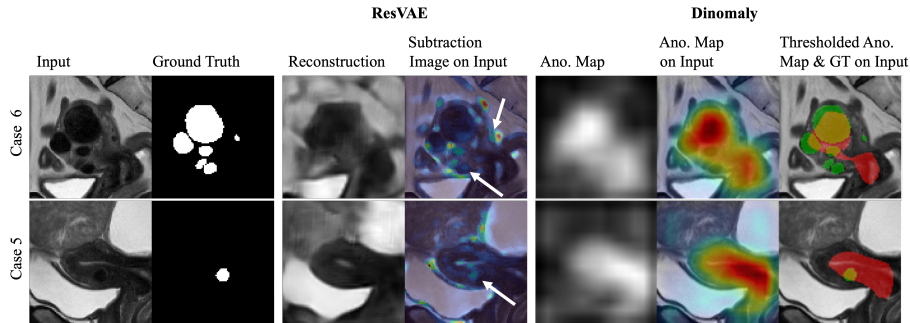


Fig. 4: Qualitative comparison on 2 representative cases: Input slice, GT mask, ResVAE reconstruction and subtraction heatmap overlay, Dinomaly anomaly map, heatmap overlay, and thresholded prediction with GT boundary (yellow: true positive, green: GT, red: false positive). Arrows indicate subtle lesions.

Table 2: Ablation study results (%) for pixel-level performance relative to the full model (Sec. 3). Scores exceeding the Base model are highlighted in bold.

Group	Configuration	AUROC $\uparrow$	Accuracy $\uparrow$	Precision $\uparrow$	Sensitivity $\uparrow$	Specificity $\uparrow$
Base	Full model	88.06	93.62	17.31	38.49	95.10
Architectural	ViT-S Backbone	83.77	90.67	16.63	42.7	92.38
	12 Decoder Layers	87.75	93.00	16.34	41.17	94.38
	8 Attention Heads	87.02	92.27	19.72	40.58	94.11
	0 Decoder MLP Ratio	88.04	93.88	17.60	36.86	95.40
Training	2 Augmentations/slice	87.23	92.97	15.98	40.04	94.11
	No Synthetic Data	87.81	93.36	16.77	39.27	94.80
	No Dropout Bottleneck	88.12	93.70	17.47	38.27	95.18
Data scaling	50% of $\mathcal{D}_{\text{train}}$	83.76	92.01	13.16	37.08	93.47

5 Discussion and Conclusion

We presented a systematic adaptation of the Dinomaly framework for unsupervised anomaly detection in pelvic T2w MRI, leveraging a frozen DINOv3 ViT-L/16 encoder with LA decoding and a noisy MLP bottleneck. The strong pixel-level AUROC (88.06%) is largely attributable to the frozen encoder, whose semantic representations generalize well to pelvic MRI and prevent co-adaptation with the decoder, preserving a stable normative reference. The ablation study confirms robustness in a data-scarce setting. At frame level, sensitivity (11.61%) is deliberately sacrificed for high specificity (95.45%), the conservative healthy-calibrated threshold minimizes false alarms, as spurious flagging would erode clinical trust. The coarse 28×28 token grid is the primary limitation, causing small or low-contrast lesions to produce insufficient anomalous responses. The complementary VAE baseline, high sensitivity (79.91%) but low specificity (64.43%), highlights the benefits of potentially fusing both paradigms. Clinically, the system is designed as a real-time decision support tool: when a frame

is flagged, the spatial anomaly map is immediately available at the scanner, enabling the radiologist to adapt. Generalization beyond myomas and Nabothian cysts, to other pathologies and interventional settings, remains to be validated.

Acknowledgement This work was supported by the Bavarian Ministry of Health, Care and Prevention [EndoKI], DFG Heisenberg [502024488], ERC StG EARTHWORM [101165242], ERC Proof-of-concept grant SYNCWORM [101293293] and CAIMed (Lower Saxony Center for Artificial Intelligence and Causal Methods in Medicine) [ZN4257] funding. M.L. receives funding from the Berlin Institute of Health (Junior Digital Clinician Scientist Grant). The authors have no competing interests to declare.

References

1. Bruno, M.A., Walker, E.A., Abujudeh, H.H.: Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *Radiographics* **35**(6), 1668–1676 (2015)
2. Guo, J., Lu, S., Zhang, W., Chen, F., Li, H., Liao, H.: Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In: CVPR. pp. 20405–20415 (2025)
3. Hernando, D., Zhang, Y., Pirasteh, A.: Quantitative diffusion mri of the abdomen and pelvis. *Medical physics* **49**(4), 2774–2793 (2022)
4. Huang, P., Shang, J., Xu, Y., Hu, Z., Zhang, K., Dai, J., Yan, H.: Anomaly detection in radiotherapy plans using deep autoencoder networks. *Frontiers in Oncology* **13**, 1142947 (2023)
5. Katharopoulos, A., Vyas, A., Pappas, N., Fleuret, F.: Transformers are rnns: Fast autoregressive transformers with linear attention. In: ICML. p. 5156. PMLR (2020)
6. Kilbride, B.F., Narsinh, K.H., Jordan, C.D., Mueller, K., Moore, T., Martin, A.J., Wilson, M.W., Hetts, S.W.: Mri-guided endovascular intervention: current methods and future potential. *Expert review of medical devices* **19**(10), 763–778 (2022)
7. Knoll, F., Hammernik, K., Zhang, C., Moeller, S., Pock, T., Sodickson, D.K., Akcakaya, M.: Deep-learning methods for parallel magnetic resonance imaging reconstruction: A survey of the current approaches, trends, and issues. *IEEE signal processing magazine* **37**(1), 128–140 (2020)
8. Knupfer, A., Müller, J.P., Verdera, J.A., Fenske, M., Mathy, C.S., Tripathy, S., Arndt, S., May, M., Uder, M., Beckmann, M.W., Burghaus, S., Hutter, J.: Unsupervised anomaly detection of diseases in the female pelvis for real-time mr imaging (2026), <https://arxiv.org/abs/2602.06179>
9. Küstner, T., Fuin, N., Hammernik, K., Bustin, A., Qi, H., Hajhosseiny, R., Masci, P.G., Neji, R., Rueckert, D., Botnar, R.M., et al.: Cinenet: deep learning-based 3d cardiac cine mri reconstruction with multi-coil complex-valued 4d spatio-temporal convolutions. *Scientific reports* **10**(1), 13710 (2020)
10. Liang, H., Tse, Z.T.H.: Mr conditional prostate intervention systems and actuations review. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* **237**(1), 18–34 (2023)
11. Meta AI: facebook/dinov3-vits16-pretrain-lvd1689m. HuggingFace Hub, <https://huggingface.co/facebook/dinov3-vits16-pretrain-lvd1689m> (2025)
12. Moynagh, M.R., Dowdy, S.C., Welch, B., Glaser, G.E., Schmitz, J.J., Jatoi, A., Langstraat, C.L., Block, M.S., Kurup, A.N., Kumar, A.: Image-guided tumor ablation in gynecologic oncology: review of interventional oncology techniques and

- case examples highlighting a collaborative, multidisciplinary program. *Gynecologic oncology* **160**(3), 835–843 (2021)
13. Müller, J.P.: Synthuterus roi (2026). <https://doi.org/10.5281/ZENODO.18297879>
 14. Pan, H., Chen, M., Bai, W., Li, B., Zhao, X., Zhang, M., Zhang, D., Li, Y., Wang, H.: UMD (2023). <https://doi.org/10.6084/m9.figshare.23541312.v3>
 15. Pan, H., Chen, M., Bai, W., Li, B., Zhao, X., Zhang, M., Zhang, D., Li, Y., Wang, H., Geng, H., Kong, W., Yin, C., Han, L., Lan, J., Zhao, T.: Large-scale uterine myoma mri dataset covering all figo types with pixel-level annotations. *Scientific Data* **11**(1) (2024). <https://doi.org/10.1038/s41597-024-03170-x>
 16. Pinaya, W.H., 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., et al.: Fast unsupervised brain anomaly detection and segmentation with diffusion models. In: MICCAI. pp. 705–714 (2022)
 17. Pötter, R., Tanderup, K., Schmid, M.P., Jürgenliemk-Schulz, I., Haie-Meder, C., Fokdal, L.U., Sturdza, A.E., Hoskin, P., Mahantshetty, U., Segedin, B., et al.: Mri-guided adaptive brachytherapy in locally advanced cervical cancer (embrace-i): a multicentre prospective cohort study. *The lancet oncology* **22**(4), 538–547 (2021)
 18. Rahmaniar, W., Suzuki, K.: Multi-ad: cross-domain unsupervised anomaly detection for medical and industrial applications. *Pattern Recognition* p. 112486 (2025)
 19. Romano, A., Votta, C., Nardini, M., Chiloire, G., Panza, G., Boldrini, L., Cusumano, D., Galofaro, E., Placidi, L., Antonelli, M.V., et al.: Uterus motion analysis for radiotherapy planning optimization. the innovative contribution of on-board hybrid mr imaging. *Clinical and Translational Radiation Oncology* **47**, 100808 (2024)
 20. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F., Haziza, D., Wehrstedt, L., Wang, J., Darcet, T., Moutakanni, T., Sentana, L., Roberts, C., Vedaldi, A., Tolan, J., Brandt, J., Couprie, C., Mairal, J., Jégou, H., Labatut, P., Bojanowski, P.: DINOv3 (2025), <https://arxiv.org/abs/2508.10104>
 21. Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J.N., Wu, Z., Ding, X.: Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical image analysis* **63**, 101693 (2020)
 22. Tripathy, S., Castro, N., May, M., Siegler, L., Story, L., Uder, M., Hutter, J.: Deep supervision attention u-net for segmentation of uterine zones: a multi-center study (2025). <https://doi.org/10.58530/2025/0105>, iSMRM 2025
 23. Uka, M., Matsui, Y., Iguchi, T., Matsushita, T., Tomita, K., Umakoshi, N., Kawabata, T., Yamada, M., Hiraki, T.: Magnetic resonance imaging guidance for percutaneous needle intervention. *Interventional Radiology* **9**(3), 92–98 (2024)
 24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *NeurIPS* **30** (2017)
 25. Wasnik, A.P., Mazza, M.B., Liu, P.S.: Normal and variant pelvic anatomy on mri. *Magnetic Resonance Imaging Clinics* **19**(3), 547–566 (2011)
 26. White, I.M., Scurr, E., Wetscherek, A., Brown, G., Sohaib, A., Nill, S., Oelfke, U., Dearnaley, D., Lalondrelle, S., Bhide, S.: Realizing the potential of magnetic resonance image guided radiotherapy in gynaecological and rectal cancer. *The British Journal of Radiology* **92**(1098), 20180670 (2019)
 27. You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., Le, X.: A unified model for multi-class anomaly detection. *NeurIPS* **35**, 4571 (2022)
 28. Yu, Y., Zhang, L., Sui, L., Zhang, H., Chen, L.: Application of mri-guided hysteroscopic one-step resection in preserving the fertility of early endometrial cancer patients. *Frontiers in Oncology* **15**, 1597185 (2025)