# Unsupervised Artifact Detection and Quantification via Contrastive Learning with Noise Reference

Levente Lippenszky[1 [0009−0000−7835−2071]], Hongxu Yang[2 [0000-0003-2447-094X]], José de Arcos[3 [0000−0002−1855−4285]], and László Ruskó[1 [0009−0003−5073−8978]]

[1] Science & Technology Org. AI & ML, GE HealthCare, Budapest, Hungary
[2] Science & Technology Org. AI & ML, GE HealthCare, Eindhoven, Netherlands
[3] Applied Science Laboratory Europe, GE HealthCare, Madrid, Spain
levente.lippenszky@gehealthcare.com

**Abstract.** Artifacts in MR images can degrade diagnostic utility and compromise the performance of downstream algorithms. Deep neural networks are particularly sensitive to such artifacts and can produce inaccurate or biased outputs. Automated artifact detection is therefore essential for improving clinical efficiency and ensuring high-quality training data. In this work, we propose a contrastive learning approach that structures the embedding space to position images with higher artifact levels closer to a noise reference. This enables unsupervised artifact detection and quantification by computing the cosine similarity between the image and noise embeddings at test time. Extensive experiments showed that our method outperforms existing unsupervised approaches in detecting various types of MR artifacts, including motion, ghosting, aliasing, metal and gas, on prostate T2-weighted and brain T1-weighted images. In addition, it achieved the highest performance in motion artifact quantification by a substantial margin, highlighting its ability to learn rich representations of artifact severity.

**Keywords:** MR artifact detection · Image quality control · Contrastive learning.

## 1 Introduction

Significant artifacts in an image can make the scan unsuitable for diagnostic applications, research, or integration into downstream algorithms that aid clinical decision-making [16]. Deep neural networks (DNNs) are known to be sensitive to image quality [9, 11], producing inaccurate or biased outputs when images are compromised by artifacts [22]. An automatic artifact detection system could improve clinical workflow efficiency by enabling immediate repetition of problematic scans and reducing the need for human image quality assessment. In addition, it could remove corrupted images from large datasets, enabling the training of more robust DNNs [16].

Current automatic image quality control (IQC) methods often rely on annotations generated by human experts [8, 3, 2, 13]. However, manual labeling is subjective, time-consuming and dataset-specific, which limits the generalization of these supervised methods. Unsupervised approaches have emerged to overcome these challenges. RadQy [18], for example, is an open-source IQC tool for MR and CT images that extracts a series of quality measures, including noise ratios, variation metrics, entropy and energy criteria. Sciarra et al. [20] trained a network to predict the structural similarity index measure between synthetically corrupted and corresponding clean images (SSIM-Regr). Zuo et al. [22] first trained an encoder to learn artifact representations using contrastive learning (CLR) and synthetically generated artifacts. Artifact detection was then performed by thresholding the likelihood of the resulting embeddings, as estimated by a normalizing flow (CLR-NF). However, the method relies on assumptions about the prevalence of artifact-corrupted images that may not generalize across datasets. Furthermore, it lacks a mechanism for quantifying artifact severity and requires training two separate models on disjoint subsets of the training data.

In this study, we propose a novel contrastive learning framework that enables both detection and quantification of MR artifacts using a single network. Specifically, we designed a contrastive loss that encourages patches with similar artifact levels to lie closer in the embedding space, using standard Gaussian noise as a reference. The contrastive loss encourages alignment between artifact-corrupted patches and noise samples, while enforcing artifact-free patches to remain dissimilar from noise. Moreover, the proposed method organizes the space such that more severely corrupted patches are embedded closer to noise than patches with less severe artifacts. Since the designed loss structures the embedding space solely through angular relationships, we defined the artifact score as the cosine similarity between the noise embedding and the image embedding at test time. The main **contributions** of this work are the following.

1. A novel contrastive learning method is proposed that structures the embedding space such that images with higher artifact levels are embedded closer to a noise reference.
2. A novel unsupervised method for artifact detection is introduced that uses a single network, without making any assumptions about artifact prevalence. In addition, it incorporates a mechanism to quantify artifact severity directly.
3. Extensive evaluations were conducted on four test sets spanning two use cases: prostate T2-weighted (T2w) and brain T1-weighted (T1w) images. Our method outperformed existing unsupervised approaches in artifact detection and achieved substantially stronger performance in artifact quantification.

## 2 Methods

### 2.1 Artifact Encoder

Contrastive learning aims to learn discriminative features by comparing query, positive, and negative examples. In this study, the positive example was selected

to have the same artifact level as the query, while negative examples were chosen to differ in artifact severity [22]. Contrastive training of the artifact encoder is illustrated in Fig. 1. We distinguished between two scenarios, A and B, each occurring with probability 0.5. In scenario A, query $\mathbf{x}^*$ and positive $\mathbf{x}^+$ were sampled as random patches from the same MR image $I$. Negative examples were constructed using two distinct strategies. In the first strategy, synthetic artifacts were introduced using a randomly selected transform from a set of artifact generators (e.g., `RandomMotion` and `RandomGhosting` from TorchIO [15]). Generator parameters were sampled uniformly from a predefined range; for example, `translation` of 3 mm could be drawn from the [2 mm, 10 mm] interval when using `RandomMotion`. Subsequently, $N_\alpha$ random patches were extracted from the artifact-corrupted version of image $I$. Artifact negatives, denoted by $\{\mathbf{x}^-_{\alpha,i}\}_{i=1}^{N_\alpha}$, guide the encoder to focus on artifact-related features while discouraging it from learning irrelevant information related to contrast or anatomy. In the second strategy, $N_\beta$ random patches were sampled from another MR image, denoted by $\{\mathbf{x}^-_{\beta,j}\}_{j=1}^{N_\beta}$. Here, we assumed that patches from a different scan exhibit different artifact level. By leveraging patches from real MR images as negatives, the encoder learns to capture various kinds of artifacts beyond those simulated during training [22]. In scenario B, we randomly selected an artifact generator $G_{\text{low}}$ to corrupt image $J$, yielding $J_{\text{low}}$. The query and positive examples were sampled as two random patches from $J_{\text{low}}$. To generate negative examples, we selected a second artifact generator $G_{\text{high}}$, constrained to produce higher artifact severity than $G_{\text{low}}$. For instance, if $G_{\text{low}}$ was a `RandomMotion` transform with a `translation` parameter of 1 mm, then $G_{\text{high}}$ could be `RandomMotion` with a 4 mm `translation`. Negative examples included $N_\alpha$ patches from high-artifact image $J_{\text{high}}$ and $N_\beta$ patches from another MR image. In each scenario, we generated $N_\nu$ noise patches from the standard normal distribution denoted by $\{\mathbf{x}_{\nu,k}\}_{k=1}^{N_\nu}$.

We trained the artifact encoder $f(\cdot)$ via contrastive learning and used noise as reference. The proposed contrastive loss $\mathcal{L}$ consists of two terms. The first term is based on the NT-Xent loss [6] and encourages patches with similar artifact levels to lie closer in the embedding space:

$$\mathcal{L}_{CLR} = -\log \frac{\exp(\text{sim}(\mathbf{z}^*, \mathbf{z}^+)/\tau)}{\exp(\text{sim}(\mathbf{z}^*, \mathbf{z}^+)/\tau) + \sum_{i=1}^{N_\alpha+N_\beta} \exp(\text{sim}(\mathbf{z}^*, \mathbf{z}_i^-)/\tau)} \tag{1}$$

where $\mathbf{z} = f(\mathbf{x}) \in \mathbb{R}^2$ is the embedding vector for input $\mathbf{x}$, $\{\mathbf{z}_i^-\}_{i=1}^{N_\alpha+N_\beta}$ are all the negative embeddings, $\text{sim}(\mathbf{u}, \mathbf{v}) = \mathbf{u}^\intercal\mathbf{v}/\|\mathbf{u}\|\|\mathbf{v}\|$ denotes the cosine similarity and $\tau$ is the temperature scaling parameter. Intuitively, $\mathcal{L}_{CLR}$ pulls the positive closer to the query and pushes negatives apart in the embedding space. The second loss term encourages alignment between artifact negatives and noise patches, while enforcing query and positive to remain dissimilar from noise samples:

$$\mathcal{L}^\nu_{CLR} = -\log \frac{\exp(\text{sim}(\bar{\mathbf{z}}_\nu, \bar{\mathbf{z}}_\alpha^-)/\tau_\nu)}{\exp(\text{sim}(\bar{\mathbf{z}}_\nu, \bar{\mathbf{z}}_\alpha^-)/\tau_\nu) + \exp(\text{sim}(\bar{\mathbf{z}}_\nu, \frac{\mathbf{z}^*+\mathbf{z}^+}{2})/\tau_\nu)} \tag{2}$$

where $\bar{\mathbf{z}}_\nu = \frac{1}{N_\nu}\sum_{i=1}^{N_\nu}\mathbf{z}_{\nu,i}$ and $\bar{\mathbf{z}}_\alpha^- = \frac{1}{N_\alpha}\sum_{j=1}^{N_\alpha}\mathbf{z}_{\alpha,j}^-$, $\tau_\nu$ denotes the temperature scaling parameter for the loss term. The artifact encoder was trained using a weighted combination of the two contrastive loss terms:

$$\mathcal{L} = \mathcal{L}_{CLR} + \lambda_\nu \mathcal{L}_{CLR}^\nu \tag{3}$$

where $\lambda_\nu$ controls the contribution of the second term. In scenario A, $\mathcal{L}$ structures the embedding space so that artifact-corrupted patches are positioned closer to noise patches than artifact-free patches. In scenario B, it arranges the space such that more severely corrupted patches are embedded closer to noise than those with less severe artifacts. Therefore, scenario A encourages the encoder to learn representations for artifact detection, while scenario B enforces learning for artifact quantification. Fig. 2(a) visualizes the embedding space, where L2-normalized embeddings lie on the unit circle.

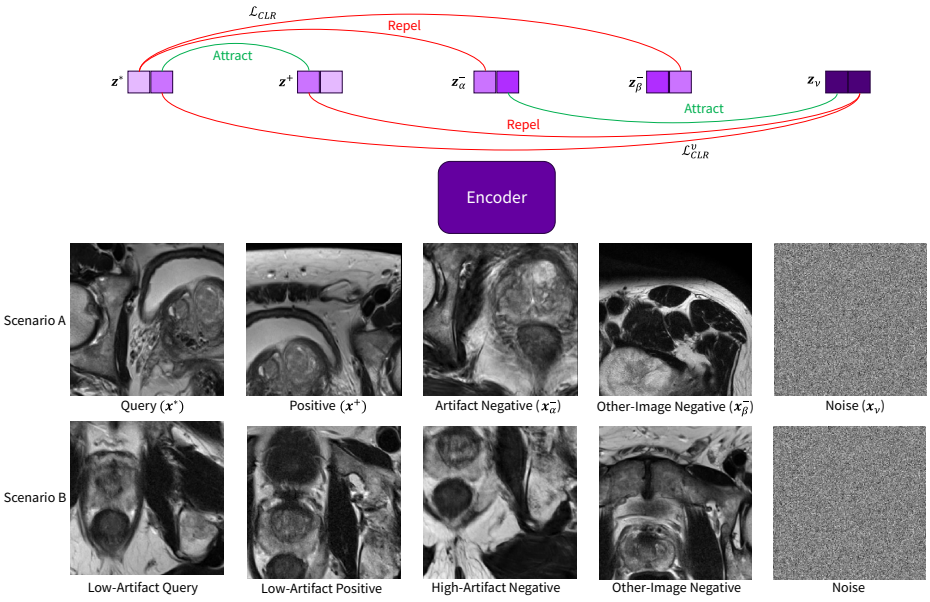
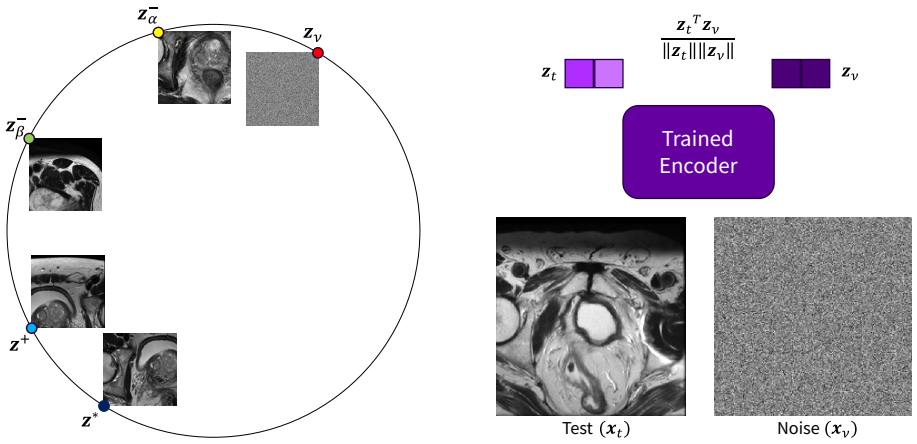



**Fig. 1.** Contrastive training of the artifact encoder. Scenarios A and B are sampled with equal probability. In A, query and positive are random patches from the same scan $I$. Artifact negative is a random patch from artifact-corrupted version of $I$. Other-image negative is sampled from another MR scan. In B, query and positive are from low-artifact version of scan $J$, high-artifact negative is from a heavily corrupted version of $J$. In embedding space, query $\mathbf{z}^*$ is attracted to positive $\mathbf{z}^+$ and repelled from negatives $\mathbf{z}_\alpha^-$ and $\mathbf{z}_\beta^-$, noise $\mathbf{z}_\nu$ is attracted to artifact negative $\mathbf{z}_\alpha^-$ and repelled from $\mathbf{z}^*$ and $\mathbf{z}^+$. During training, multiple negatives and noise patches are sampled per query. In scenario A, artifact-corrupted patches are positioned closer to noise than artifact-free ones (artifact detection); in B, patches with high artifact levels are embedded closer to noise than patches with low artifact levels (artifact quantification).

### 2.2 Artifact Score

Since the loss is based on the cosine similarity, the embedding space is structured solely by angular relationships. At test time, noise patches are generated from the standard normal distribution and a center crop is extracted from the image. We defined the artifact score as the cosine similarity between the average noise embedding and the embedding of the center crop, as shown in Fig. 2(b) using a single noise patch. Higher score indicates higher artifact level.




(a) Embedding space. (b) Artifact score.

**Fig. 2.** (a) L2-normalized embeddings lie on the unit circle in $\mathbb{R}^2$. Artifact negative $\mathbf{z}_\alpha^-$ is embedded closer to noise $\mathbf{z}_\nu$ than query $\mathbf{z}^*$ and positive $\mathbf{z}^+$. (b) Artifact score is defined as the cosine similarity between noise and test image embeddings.

## 3 Experimental Results

### 3.1 Datasets

For the prostate T2w use case, the training set comprised 720 and the validation set 180 scans from the public PI-CAI dataset [19]. We evaluated the methods on one real and two synthetic test sets. We annotated a private dataset of 177 real-world prostate T2w scans for MR artifacts (Real Multi). The dataset contained retrospectively collected, anonymized clinical data; IRB approval was not required for secondary use. It included 11 motion, 18 ghosting, 11 metal, 3 aliasing and 1 gas artifact cases. We also generated a synthetic multiclass test set (Synth Multi), in which 600 held-out cases from the PI-CAI dataset constituted the artifact-free class. For each artifact class, 600 cases were simulated from the artifact-free scans. Motion and ghosting were generated using `RandomMotion`

and `RandomGhosting` transforms, while aliasing was simulated following the implementation in [16]. Furthermore, we created a synthetic motion-specific test set (Synth Motion), consisting of moderate and severe motion artifact classes, each simulated from the 600 artifact-free cases. Scans were first preprocessed by standardizing their orientation to RAS+, followed by resampling to an isotropic resolution of $0.6 \times 0.6 \times 0.6$ mm$^3$. Image intensities were rescaled to $[-1, 1]$ using the $[0.25, 99.75]$ percentiles of each input image, reducing the influence of extreme intensity values, such as those caused by hip implants in prostate scans [4]. We used a patch size of $180 \times 180 \times 70$ to extract query, positive and negative examples.

For the brain T1w use case, the training and validation sets consisted of 523 and 58 scans from the public IXI dataset [1], respectively. We utilized the public MR-ART dataset [14] as test set. The dataset contains 436 T1w brain scans from neurologically healthy adults, each acquired under three conditions: completely still, mild head motion, and pronounced head motion. Three neuroradiologists assigned each volume a 3-point motion artifact severity score. We preserved the original near-isotropic spacing of the IXI dataset ($0.9 \times 0.9 \times 1.2$ mm$^3$), and used this resolution during evaluations. After rescaling image intensities to $[-1, 1]$, we extracted $128 \times 128 \times 74$ patches for contrastive training.

### 3.2 Implementation Details and Comparison Methods

In both use cases, a mini-batch contained one query, one positive, ten artifact negatives ($N_\alpha = 10$), ten other-image negatives ($N_\beta = 10$) and two noise patches generated from the standard Gaussian distribution ($N_\nu = 2$). We set the temperature scaling parameters as $\tau = 0.1$ and $\tau_\nu = 1$. By setting $\tau < \tau_\nu$, the contrastive loss emphasizes stronger attraction between the query and positive than between artifact negatives and noise samples. We trained a DenseNet121 [10] model, as implemented in MONAI 1.3.2 [5], using the Adam optimizer [12] with a learning rate of $10^{-3}$. We applied `RandomMotion` with `translation` parameter [2 mm, 10 mm] and `RandomGhosting` with `num_ghosts` parameter $[4, 10]$ to generate artifact negative examples. We set $\lambda_\nu = 0.5$, which controls the contribution of $\mathcal{L}^{\nu}_{CLR}$ in the loss. The artifact encoder was trained for 200 epochs for the prostate T2w and 300 epochs for the brain T1w use case. Model weights were saved at the lowest validation loss. Trainings were performed on an NVIDIA GeForce RTX 3090 GPU card with 24 GB memory.

As no public code was available, we re-implemented CLR-NF [22]. For the prostate T2w use case, we trained the artifact encoder on 150 PI-CAI cases. RealNVP [7] normalizing flow was trained on artifact embeddings generated on 750 PI-CAI images using the glasflow library [21]. For the brain T1w use case, the artifact encoder and RealNVP were trained on 100 and 481 IXI scans, respectively. We also re-implemented SSIM-Regr [20]. Lastly, we evaluated RadQy and reported the results of its best-performing metric among the 15 provided. For a fair comparison, we applied the same settings as for our method wherever applicable.

### 3.3 Artifact Detection and Quantification

Table 1 presents the artifact detection performance of all methods across the four test sets. Each row reports the binary classification performance for distinguishing the artifact-free class (label 0) from the corresponding artifact class (label 1), measured by the area under the receiver operating characteristic curve (AUROC). On the MR-ART dataset, our method outperforms all other approaches for both moderate and severe motion artifacts. On the Real Multi test set, it achieves superior performance across all artifact types except for ghosting. In Cartesian sampling schemes, ghosting artifacts appear along the phase-encoding direction and are more prominent near the edges of the image, where signal intensity is lower [17]. All methods, except RadQy, apply center cropping during inference, potentially excluding artifact-affected regions. Increasing inference patch size of our method to $360 \times 360 \times 140$ improved ghosting detection to an AUROC of 0.83, supporting this hypothesis. Overall, our method demonstrates the strongest performance on synthetic test sets. Fig. 3 shows axial slices from artifact-affected scans that were correctly detected by our method.

**Table 1.** Artifact detection performance of our method compared against RadQy [18], SSIM-Regr [20] and CLR-NF [22] on four test sets. Each row reports the binary classification performance for distinguishing between the artifact-free class (label 0) and the artifact class (label 1) in AUROC. 95% bootstrap CI widths are shown in brackets. Best performance is highlighted in **bold**. Underlining indicates that our method performs statistically significantly better than the second-best method at a 5% level. Statistical test was conducted using 95% bootstrap CI for the AUROC difference, results were considered significant if the CI did not include zero.

| Test Set | Artifact | RadQy | SSIM-Regr | CLR-NF | CLR (Ours) |
|---|---|---|---|---|---|
| MR-ART (T1w) | Mod Motion | 0.62 (0.13) | 0.55 (0.14) | 0.53 (0.14) | **0.68** (0.14) |
| | Sev Motion | 0.65 (0.12) | 0.63 (0.12) | 0.61 (0.12) | **0.94** (0.05) |
| Real Multi (T2w) | Motion | 0.65 (0.38) | 0.61 (0.37) | 0.73 (0.28) | **0.79** (0.19) |
| | Ghosting | **0.86** (0.14) | 0.84 (0.20) | 0.55 (0.26) | 0.27 (0.18) |
| | Metal | 0.78 (0.28) | 0.66 (0.30) | 0.50 (0.35) | **0.97** (0.05) |
| | Aliasing | 0.78 (0.45) | 0.08 (0.16) | **0.98** (0.05) | **0.98** (0.04) |
| | Gas | 0.00 (0.00) | 0.45 (0.17) | 0.71 (0.15) | **0.72** (0.15) |
| Synth Multi (T2w) | Motion | 0.81 (0.05) | **1.00** (0.00) | 0.92 (0.03) | **1.00** (0.01) |
| | Ghosting | 0.58 (0.06) | 0.92 (0.03) | 0.59 (0.06) | **0.93** (0.03) |
| | Aliasing | 0.59 (0.06) | 0.99 (0.01) | 0.80 (0.05) | **1.00** (0.00) |
| Synth Motion (T2w) | Mod Motion | 0.82 (0.05) | **1.00** (0.00) | 0.93 (0.03) | **1.00** (0.01) |
| | Sev Motion | 0.83 (0.05) | **1.00** (0.00) | 0.95 (0.02) | **1.00** (0.00) |

We evaluated the methods for artifact quantification by computing the Spearman correlation between ground truth labels and artifact scores on the MR-ART and synthetic motion test sets. Table 2 demonstrates that our method significantly outperforms other approaches on both datasets.

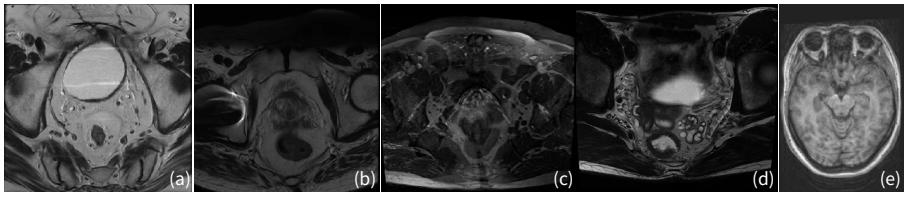



**Fig. 3.** Examples of artifacts from the Real Multi (a–d) and MR-ART (e) test sets, correctly detected by our method. The artifact types shown are: (a) ghosting, (b) metal, (c) aliasing, (d) gas and (e) motion.

**Table 2.** Spearman correlation between predicted artifact scores and ground truth labels for each method, evaluated on two test sets.

| Test Set | Artifact | RadQy | SSIM-Regr | CLR-NF | CLR (Ours) |
|---|---|---|---|---|---|
| MR-ART (T1w) | Motion | 0.23 (0.19) | 0.21 (0.18) | 0.17 (0.18) | **0.72** (0.09) |
| Synth Motion (T2w) | Motion | 0.48 (0.08) | 0.74 (0.05) | 0.66 (0.06) | **0.91** (0.01) |

### 3.4 Ablation studies

To assess the contribution of individual components of our method, we conducted ablation studies on the MR-ART dataset. First, we did not include any other-image negatives during contrastive training ($N_\beta = 0$). Second, we removed the loss term that structures the embedding space to position noise as reference ($\lambda_\nu = 0$). Finally, we excluded scenario B, where the query and positive examples contain low-severity artifacts. AUROC in Table 3 quantifies performance in distinguishing artifact-free scans from all motion-affected cases. The results demonstrate that including all components yields the best performance.

**Table 3.** Ablation studies for individual components of our method conducted on the MR-ART test set. $N_\beta$: number of other-image negatives, $\lambda_\nu$: weight for the second loss term.

| Method | AUROC | Spearman Corr |
|---|---|---|
| CLR (Ours) | **0.85** (0.07) | **0.72** (0.09) |
| $N_\beta = 0$ | 0.70 (0.11) | 0.43 (0.17) |
| $\lambda_\nu = 0$ | 0.66 (0.10) | 0.41 (0.17) |
| No Scenario B | 0.71 (0.12) | 0.39 (0.17) |

## 4 Conclusion

In this work, we proposed a contrastive learning method that structures the embedding space such that images with higher artifact levels lie closer to a noise reference. This enables unsupervised detection and quantification of various MR

artifacts in both prostate T2w and brain T1w images. Our method outperformed existing unsupervised approaches in artifact detection and achieved the highest performance in artifact quantification by a substantial margin, highlighting its ability to learn rich representations of artifact severity. Since our method does not rely on manual annotations, it enables the training of large-scale models on a diverse range of unannotated datasets. Future work could explore training a foundation model on multi-sequence, multi-anatomy MR datasets to enable generalization across diverse use cases. Given the strong artifact detection performance achieved with a lightweight architecture, our method enables fast inference and could be integrated into clinical workflows to flag low-quality scans for reacquisition or downstream correction.

**Acknowledgments.** This study was funded under the 2023 Call for "R&D Projects linked to Personalized Medicine and Advanced Therapies within the framework of the ISCIII–CDTI Joint Initiative", financed through the "Recovery, Transformation and Resilience Plan – Funded by the European Union – NextGenerationEU".

**Disclosure of Interests.** This work was conducted by the authors as part of their full-time employment at GE HealthCare.

## References

1. IXI dataset, https://brain-development.org/ixi-dataset/, data made available under the Creative Commons CC BY-SA 3.0 license.
2. Ahmad, A., Parker, D., Dheer, S., Samani, Z.R., Verma, R.: 3D-QCNet–A pipeline for automated artifact detection in diffusion MRI images. Computerized Medical Imaging and Graphics **103**, 102151 (2023)
3. Belue, M.J., Law, Y.M., Marko, J., Turkbey, E., Malayeri, A., Yilmaz, E.C., Lin, Y., Johnson, L., Merriman, K.M., Lay, N.S., et al.: Deep learning-based interpretable AI for prostate T2W MRI quality evaluation. Academic Radiology **31**(4), 1429–1437 (2024)
4. Breit, H.C., Vosshenrich, J., Clauss, M., Weikert, T.J., Stieltjes, B., Kovacs, B.K., Bach, M., Harder, D.: Visual and quantitative assessment of hip implant-related metal artifacts at low field MRI: a phantom study comparing a 0.55-T system with 1.5-T and 3-T systems. European Radiology Experimental **7**(1), 5 (2023)
5. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
6. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PmLR (2020)
7. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using Real NVP. arXiv preprint arXiv:1605.08803 (2016)
8. Esteban, O., Birman, D., Schaer, M., Koyejo, O.O., Poldrack, R.A., Gorgolewski, K.J.: MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. PloS one **12**(9), e0184661 (2017)
9. Hendrycks, D., Dietterich, T.: Benchmarking neural network robustness to common corruptions and perturbations. arXiv preprint arXiv:1903.12261 (2019)

10. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
11. Jaspers, T.J., Boers, T.G., Kusters, C.H., Jong, M.R., Jukema, J.B., de Groof, A.J., Bergman, J.J., de With, P.H., van der Sommen, F.: Robustness evaluation of deep neural networks for endoscopic image analysis: Insights and strategies. Medical Image Analysis **94**, 103157 (2024)
12. Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
13. Lim, A., Lo, J., Wagner, M.W., Ertl-Wagner, B., Sussman, D.: Automatic artifact detection algorithm in fetal MRI. Frontiers in Artificial Intelligence **5**, 861791 (2022)
14. Ádám Nárai, Hermann, P., Auer, T., Kemenczky, P., Szalma, J., Homolya, I., Somogyi, E., Vakli, P., Weiss, B., Vidnyánszky, Z.: "Movement-related artefacts (MR-ART) dataset" (2022). https://doi.org/doi:10.18112/openneuro.ds004173.v1.0.2
15. Pérez-García, F., Sparks, R., Ourselin, S.: TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Computer methods and programs in biomedicine **208**, 106236 (2021)
16. Ravi, D., Barkhof, F., Alexander, D.C., Puglisi, L., Parker, G.J., Eshaghi, A., Initiative, A.D.N., et al.: An efficient semi-supervised quality control system trained using physics-based MRI-artefact generators and adversarial training. Medical Image Analysis **91**, 103033 (2024)
17. Runge, V.M., Nitz, W.R., Schmeets, S.H., Faulkner, W.H., Desai, N.K.: The physics of clinical MR taught through images. Springer (2009)
18. Sadri, A.R., Janowczyk, A., Zhou, R., Verma, R., Beig, N., Antunes, J., Madabhushi, A., Tiwari, P., Viswanath, S.E.: MRQy—An open-source tool for quality control of MR imaging data. Medical physics **47**(12), 6029–6038 (2020)
19. Saha, A., Twilt, J.J., Bosma, J.S., van Ginneken, B., Yakar, D., Elschot, M., Veltman, J., Fütterer, J., de Rooij, M., Huisman, H.: The PI-CAI Challenge: Public Training and Development Dataset (Jun 2022). https://doi.org/10.5281/zenodo.6624726, https://doi.org/10.5281/zenodo.6624726
20. Sciarra, A., Chatterjee, S., Dünnwald, M., Placidi, G., Nürnberger, A., Speck, O., Oeltze-Jafra, S.: Reference-less SSIM regression for detection and quantification of motion artefacts in brain MRIs. In: Medical imaging with deep learning (2022)
21. Williams, M.J., jmcginn, federicostak, Veitch, J.: uofgravity/glasflow: v0.4.1 (Oct 2024). https://doi.org/10.5281/zenodo.13914483, https://doi.org/10.5281/zenodo.13914483
22. Zuo, L., Xue, Y., Dewey, B.E., Liu, Y., Prince, J.L., Carass, A.: A latent space for unsupervised MR image quality control via artifact assessment. In: Medical Imaging 2023: Image Processing. vol. 12464, pp. 286–291. SPIE (2023)