Quality Assessment, Enhancement, and Subcortical Segmentation of Ultra-Low-Field Pediatric Brain MRI: The LISA 2026 Challenge

Yingqi Hao*, Peiwen Sui*, Mingxuan Liu, Xihai Zhao†, and Qiyuan Tian†,‡

School of Biomedical Engineering, Tsinghua Medicine, Tsinghua University

Abstract. Ultra-low-field (0.064 T) MRI offers a portable and affordable route to pediatric neuroimaging in low-resource settings, but its reduced signal-to-noise ratio, low tissue contrast, and frequent artifacts make automated analysis difficult. We present our submissions to all three tasks of the MICCAI 2026 LISA challenge. For **Task 1a** (image quality assessment), we cast the problem as multi-label ordinal classification over seven artifact categories and use a frozen Vision Transformer feature extractor trained slice-wise, followed by scan-level probability pooling and per-artifact ordinal-threshold calibration. For **Task 1b** (image quality enhancement), we adapt an Unpaired Neural Schrödinger Bridge (UNSB) to translate artifacted images into the artifact-free domain without paired ground truth. For **Task 2** (subcortical segmentation), we integrate a SwinUNETR with Euclidean distance-transform supervision, bilateral consistency, and a five-fold ensemble with sagittal test-time augmentation and light morphological post-processing. We report validation-set results and discuss how these established components were adapted to the ultra-low-field challenge setting.

Keywords: Low-field MRI · Pediatric brain · Quality assessment · Image enhancement · Schrödinger bridge · Segmentation.

1 Introduction

Accurately characterizing the developing brain in early life is central to the study of healthy neurodevelopment and to the early detection of neurodevelopmental disorders. In low- and middle-income countries, however, high-field (1.5 T/3 T) MRI is often inaccessible due to cost, infrastructure, and the need for sedation in young children. Portable ultra-low-field systems such as the 0.064 T Hyperfine SWOOP scanner have emerged as a practical alternative [1,3], trading image quality for portability, low cost, and sedation-free operation. The resulting degradation—low signal-to-noise ratio, poor gray/white matter contrast, and a range of acquisition artifacts—nonetheless makes automated quality control, enhancement, and structural segmentation challenging. Recent data-free denoising and diffusion-model optimization strategies, such as Noise2Average and DI-

* Equal contribution. † Co-senior authors. ‡ Corresponding author.

MOND, respectively, further reflect the need for specialized approaches when clean targets or stable acquisition conditions are unavailable [13,12].

The MICCAI 2026 LISA (Low-field pediatric brain Image Segmentation and quality Assurance) challenge [16,15] targets this setting with three tasks: (i) Task 1a, quality assessment via multi-label ordinal scoring of seven artifact categories; (ii) Task 1b, unpaired quality enhancement that suppresses noise and motion artifacts; and (iii) Task 2, segmentation of eleven subcortical and midline structures. We participated in all three tasks. This short paper describes our methods and validation-set results, with an emphasis on practical adaptation rather than architectural novelty. Across the tasks, we use the available structure of the problem—ordinal severity, unpaired domains, and bilateral anatomy—to guide the integration of established models under the small-data and low-quality conditions of ultra-low-field MRI.

The contribution of this challenge report is therefore the task-specific adaptation, integration, and validation of established components, together with transparent reporting of design choices and negative results. In particular:

- **Task 1a.** A lightweight quality-assessment pipeline that adapts a frozen Vision Transformer [4] and uses scan-level pooling with per-artifact ordinal-threshold calibration at the decision layer.
- **Task 1b.** An adaptation of the Unpaired Neural Schrödinger Bridge [10] to the challenge's unpaired artifact-removal setting, with a three-step configuration selected from a perceptual–fidelity trade-off.
- **Task 2.** An integrated segmentation pipeline combining SwinUNETR [6], distance-transform and bilateral supervision motivated by prior work [20], and practical ensemble, test-time augmentation, and post-processing choices.

2 Task 1a: Quality Assessment

Task 1a requires predicting, for each ultra-low-field scan, a severity score in $\{0, 1, 2\}$ for each of seven artifact categories: *Noise*, *Zipper*, *Positioning*, *Banding*, *Motion*, *Contrast*, and *Distortion*. This is a multi-label, multi-class ordinal problem with strong class imbalance (roughly 80% of labels are severity 0). Submissions are ranked by a *composite* score: the mean of five quantities computed over the flattened set of artifact–severity predictions—accuracy and support-weighted precision, recall, F1, and F2. Because the metric flattens all artifacts together and weights by support, the minority (mild/severe) severities dominate the difficulty while contributing little support, and each artifact's decision boundary influences the global weighted score.

This formulation is related to the broader no-reference quality-assessment setting, where image content is used to estimate quality without a pristine reference; for example, METER uses a multi-task transformer for no-reference quality assessment [31]. In the fetal-brain setting, label-free quality control has also been studied through uncertainty associated with image orientation recognition [17]. These works motivate quality-aware analysis, but our Task 1a target remains the challenge-defined seven-category ordinal artifact labels.

Our base model adapts the LISA 2025 winning approach for this task [16]: a frozen ImageNet-pretrained Vision Transformer [4] with slice-wise training and scan-level pooling of slice probabilities. This provides a strong, low-variance baseline on the small challenge set. Attempts to improve the representation itself (partial fine-tuning, end-to-end training, and domain-specific backbones such as a biomedical CLIP encoder) did not consistently surpass this baseline in our held-out experiments. Thus, under the present dataset size and evaluation protocol, changes at the decision layer appeared more reliable than changes to the representation; this is an empirical observation rather than a general claim about quality assessment. We therefore target the decision layer with an error-driven, per-artifact calibration.

2.1 Method

Backbone and out-of-fold probabilities. Each scan is a single-orientation (axial, coronal, or sagittal) 3D T2 volume. We extract 2D slices along the axis with the largest voxel spacing (the through-plane direction), and normalize each slice by clipping to its 1st–99th foreground percentiles followed by z-scoring. Slices are rescaled to $[0, 1]$, resized with aspect ratio preserved so that the longer side is 224 voxels, and zero-padded to 224×224. The backbone is a frozen ImageNet-pretrained ViT-B/16; only a single linear head is trained, mapping the 768-dimensional class token to 7×3 logits (seven artifacts $\times$ three severities). Each slice inherits the scan-level label, so training is weakly supervised at the slice level, and per-scan class probabilities are obtained by pooling slice probabilities (mean, maximum, and top-k averaging). Using a *subject-level* five-fold split—so no subject's orientations are split across folds—we generate out-of-fold (OOF) probabilities, such that every scan is scored by a model that did not train on it. All calibration operates on these OOF probabilities, which avoids tuning directly on predictions from the model that produced them.

Per-artifact error decomposition. For each artifact and severity level we decompose OOF errors into false negatives (missed positives; low recall) and false positives (false alarms; low precision) and label the dominant error type. This yields an interpretable map of where the composite loss originates and in which direction each artifact should be corrected. In our data the losses are strongly asymmetric: several minority severities are dominated by false negatives (under-detection), while a few severe classes are dominated by false positives (over-detection).

Ordinal-threshold rule family. Each artifact decision uses an ordinal-threshold rule on the pooled probabilities: predict severity 2 when $P(y{=}2)$ exceeds a threshold t_2; else predict severity 1 when $P(y{\geq}1)$ exceeds a threshold t_1; else predict 0. Candidate rules are generated by a grid search over (t_1, t_2), over the three pooling operators, and under several selection objectives: the composite itself, recall-oriented objectives (weighted recall, weighted F2) for under-detected artifacts, and precision-oriented objectives for over-detected artifacts. Two regularizers control small-sample variance: *threshold shrinkage* toward a neutral 0.5

boundary, and minimum-support guards that fall back to the neutral threshold when positive counts are too small to estimate a reliable cut.

Greedy global-composite integration. Because the evaluation metric flattens artifacts and weights by support, per-artifact rules cannot be chosen independently on each artifact's own score. We integrate them by greedy coordinate ascent on the *global* composite: starting from a per-artifact composite-optimal configuration, we repeatedly re-select each artifact's rule to maximize the global composite with the other six artifacts held fixed, iterating until no artifact changes. This lets an artifact adopt a different objective (e.g. a recall-oriented rule for an under-detected artifact) only when doing so improves the global score.

Honest estimation and acceptance gate. Selecting rules and reporting on the same OOF rows is optimistic. We therefore estimate performance with *nested* cross-validation: within each outer fold, the full selection is fit on the remaining folds and applied to the held-out fold; the concatenated held-out predictions give a nested estimate designed to reduce selection optimism. A candidate is accepted only if (i) its nested composite exceeds the incumbent and (ii) a subject-clustered bootstrap (resampling subjects, 5000 replicates) yields $P(\Delta{>}0) > 0.75$. This discipline rejected several representation-level methods in the available experiments and reduces the risk of shipping regressions.

Implementation details. The linear head uses AdamW [18] for 30 epochs (learning rate 10^{-3}, weight decay 10^{-4}), with a cosine-annealing schedule, automatic mixed precision, batch size 32, and 32 sampled slices per scan; five-fold models are ensembled for submission. Online augmentation is geometric only (random flips, small affine shifts, and rotations). We additionally explored ordinal training objectives—an Earth Mover's Distance (squared-CDF) loss [8], focal loss [14], and rank-consistent CORN ordinal regression [27]—together with attention-based multiple- instance learning [9] and heterogeneous probability-level ensembling. These alternatives modify the training objective, slice aggregation, or probability combination, whereas our calibration modifies the post-hoc decision rule. In the available experiments, none consistently beat the frozen-backbone-plus-calibration line under the nested acceptance gate. We do not claim that this establishes a universal advantage over the alternatives, because their detailed quantitative comparisons are not included here.

2.2 Results

Table 1 reports support-weighted quality-assessment metrics on the official validation set, together with the challenge composite (the mean of accuracy and weighted precision, recall, F1, and F2). Per-artifact error-driven calibration improves the composite over raw argmax decoding from 0.796 to 0.820; the macro-F1 also rises from 0.488 to 0.567. These results support the value of decision-layer calibration for this validation protocol, but do not by themselves establish that it is superior to every alternative training objective.

Table 1. Task 1a official validation-set results (support-weighted metrics; higher is better). Composite is the challenge ranking metric: the mean of accuracy and weighted precision, recall, F1, and F2.

Method	Accuracy	Precision	Recall	F1	F2	Composite
Raw argmax	0.812	0.798	0.812	0.766	0.790	0.796
Ordinal calibration (ours)	**0.833**	**0.823**	**0.833**	**0.797**	**0.815**	**0.820**

Table 2. Numerical Task 1a ablation summary. The all-OOF estimates are optimistic; nested estimates are used for selection.

Variant	**Composite**
Max-pool argmax baseline (all-OOF)	0.8199
Per-artifact calibration (all-OOF, optimistic)	0.8375
Per-artifact composite-optimal calibration (nested)	0.8168
Greedy global-composite integration (nested)	0.8205 ($\Delta = +0.0037$; 95% CI $[0.0004, 0.0069]$; $P(\Delta > 0) = 0.9884$)
Official validation: per-artifact calibration → global integration	0.831 → 0.833

The nested comparison isolates the benefit of optimizing the flattened challenge metric globally rather than selecting each artifact independently; the all-OOF values are shown only as optimistic references. The official transfer from 0.831 to 0.833 provides a consistent validation-set check of this integration choice.

3 Task 1b: Quality Enhancement

Task 1b asks for an unsupervised method that transforms artifacted ultra-low-field scans (affected by noise and/or motion) into images consistent with the artifact-free scans, without pixel-aligned ground-truth pairs or high-field references. Quality is measured by six no-reference and distributional metrics: Fréchet Inception Distance (FID) [7], Learned Perceptual Image Patch Similarity (LPIPS) [29], peak signal-to-noise ratio (PSNR), the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [21], CLIP-IQA [28], and Fréchet Radiomics Distance (FRD) [11]. Recent MRI restoration work has explored cycle-conditional diffusion for correction from unpaired data [30] and conditional diffusion for ultra-low-field MRI enhancement [23]. Related diffusion-based imaging workflows have also considered deformation-driven contrast-enhanced MRI synthesis [25], contrast amplification for dose-reduced brain MRI [24], and cascaded diffusion with segment-anything models for medical image synthesis [22]. These developments provide context for our task-specific adaptation of UNSB to the LISA enhancement setting.

3.1 Method

Unpaired Neural Schrödinger Bridge. We adopt UNSB [10], a GAN-based multi-step formulation that learns an entropic optimal-transport bridge between the degraded and artifact-free domains and generates the enhanced image through stochastic refinement. Building on contrastive unpaired translation [26], UNSB uses four networks: a ResNet generator G conditioned on time and noise; a PatchGAN discriminator D with time conditioning and least-squares loss [19]; a potential/energy network E for the Schrödinger-bridge entropy term; and a projection head F for the patchwise contrastive (PatchNCE) loss. The total generator objective combines adversarial, bridge, and multilayer PatchNCE terms; the latter is used to preserve anatomical content between input and output.

Domain setup and preprocessing. Using the challenge-provided categorization, we define the target (clean) domain as scans labeled *No Noise + No Motion*, and the source (degraded) domain as the union of the three remaining groups (with noise and/or motion); the generator is trained in the degraded→clean direction. We operate on 2D slices taken along the largest-spacing axis, normalized per slice by 1st–99th percentile clipping to $[-1, 1]$, with random 192×192 crops during training. Training uses unpaired sampling from the two domains (no correspondence between source and target slices).

Number of refinement steps. The number of function evaluations (refinement steps) T trades distributional and perceptual realism against pixel-level fidelity. We compared $T \in \{3, 5\}$ and adopt the $T = 3$ ("T3") configuration. T3 markedly lowers FID (155.2 → 137.8) and improves CLIP-IQA (0.301 → 0.430) and LPIPS, indicating closer distributional and perceptual agreement with the clean domain. This choice is not uniformly optimal: T5 performs better on PSNR, BRISQUE, and FRD, as reported in Table 3. We selected T3 because our submission prioritized the perceptual and distributional aspects of mapping artifacted images toward the artifact-free domain. The selection should therefore be understood as a task-specific trade-off, rather than as evidence that shorter bridges are preferable for every image-quality objective.

Implementation details. Networks are trained with Adam (learning rate 2×10^{-4}, $\beta = (0.5, 0.999)$) using a linear learning-rate schedule and least-squares adversarial loss, with loss weights $\lambda_{\text{GAN}} = \lambda_{\text{NCE}} = \lambda_{\text{SB}} = 1$ and Schrödinger-bridge noise scale $\tau = 0.01$. At inference, each validation volume is processed slice-by-slice with the $T = 3$ bridge sampler; enhanced slices are de-normalized back to the original intensity range and restacked into the native volume geometry.

3.2 Results

Table 3 reports the six challenge metrics on the validation set for our ablation over the bridge length T. The shorter $T = 3$ bridge attains the best FID, LPIPS, and CLIP-IQA, whereas the longer $T = 5$ bridge is better on PSNR, BRISQUE,

Table 3. Task 1b validation-set results (mean over cases). FID, LPIPS, BRISQUE, and FRD: lower is better; PSNR and CLIP-IQA: higher is better.

Bridge	FID ↓	LPIPS ↓	PSNR ↑	BRISQUE ↓	CLIP-IQA ↑	FRD ↓
UNSB (T=5)	155.21	0.469	**12.23**	**63.30**	0.301	**7.20**
UNSB (T=3, ours)	**137.75**	**0.463**	11.99	69.00	**0.430**	9.16

and FRD. In this unpaired setting, these metrics capture different notions of quality: distributional and perceptual agreement with the clean domain need not coincide with pixel-level or radiomics fidelity. We therefore submit T3 as a deliberate trade-off, while acknowledging that it gives lower pixel- and radiomics-based scores on the reported validation set. A recent comprehensive evaluation of unsupervised enhancement for volumetric fetal-brain MRI similarly highlights the importance of comparing complementary quality criteria rather than relying on a single score [5].

4 Task 2: Subcortical Segmentation

Task 2 requires segmenting eleven foreground labels from ultra-low-field T2 volumes: bilateral hippocampi, lateral ventricles, caudate nuclei, putamen+globus pallidus, and thalami, plus the midline corpus callosum. Ten of the eleven labels form left/right bilateral pairs, a structure we exploit explicitly. Performance is measured by Dice similarity coefficient (DSC), Hausdorff distance (HD), 95th percentile HD (HD95), average symmetric surface distance (ASSD), and relative volume error (RVE).

4.1 Method

Symmetry-aware SwinUNETR. Our pipeline uses a SwinUNETR [6] backbone (implemented with MONAI [2]) with two output heads: a segmentation head producing twelve-channel logits (eleven structures plus background) and an auxiliary head that regresses a per-class normalized Euclidean distance transform (EDT). Requiring the network to also predict each label's distance map encourages it to learn boundary geometry, which is helpful given the weak edges of ultra-low-field images. This design follows the EDT- and symmetry-based supervision introduced by Marinov et al. [20]; our contribution here is its integration with the selected SwinUNETR training and inference pipeline for this challenge, rather than a new segmentation architecture.

Training objective. The total loss combines (i) a Dice + cross-entropy segmentation loss, (ii) an L_1 loss on the EDT prediction (weight 0.5), and (iii) a *bilateral consistency* loss (weight 0.2). The consistency term flips the volume along the sagittal axis, swaps the left↔right label channels of the predicted softmax,

and minimizes the symmetric KL divergence between the original and flipped-and-swapped predictions. This enforces anatomical mirror consistency between paired structures even where boundaries are barely visible at 0.064 T. Because symmetry is handled in the loss, we deliberately do *not* use left/right flip augmentation during training.

Preprocessing and augmentation. Volumes are reoriented to RAS and intensity-normalized by 1st–99th percentile scaling to $[0, 1]$ (the challenge data are already at $1\,\mathrm{mm}^3$ isotropic resolution). Training uses foreground-biased random 96^3 patch cropping. Augmentation is intentionally conservative—small random affine transforms, random bias field, Gaussian noise, and contrast adjustment—and excludes left/right flips.

Ensembling, test-time augmentation, and post-processing. We train a five-fold cross-validation ensemble and select the best checkpoint per fold by validation DSC. At inference we use sliding-window prediction (96^3 window, 0.5 overlap, Gaussian blending). We apply *sagittal-only* two-way flip test-time augmentation with left↔right channel swap; because the network is regularized toward this symmetry during training, the flipped pass is in-distribution and complements the original prediction. Softmax probabilities are averaged across the two flip passes and the five folds before taking the argmax. Finally, per-class post-processing applies a single-iteration boundary-honest binary closing followed by largest-connected-component selection to remove spurious fragments while preserving neighboring structures.

Implementation details. The model is trained for 300 epochs with AdamW (learning rate 10^{-4}, weight decay 10^{-5}), cosine-annealing schedule, mixed precision, and batch size 6 (with two crops sampled per volume). We found that widening the sliding-window overlap beyond 0.5 did not help and could slightly regress boundary metrics, so 0.5 was retained for the final submission.

4.2 Results

Table 4 reports segmentation metrics on the official validation set, averaged over the eleven foreground structures (± standard deviation across cases). Our final pipeline—the five-fold ensemble with sagittal test-time augmentation and light morphological post-processing—attains a mean DSC of 0.82 with low boundary error (HD95 1.83, ASSD 0.50) and a relative volume error of 0.13.

5 Conclusion

We described our submissions to all three tasks of the MICCAI 2026 LISA challenge for ultra-low-field pediatric brain MRI. The main contribution is a practical, task-aware integration of established components under the constraints of small datasets and low image quality, rather than a new architecture for

Table 4. Task 2 official validation-set results, averaged over the eleven foreground structures (± std. across cases). DSC: higher is better; HD/HD95/ASSD/RVE: lower is better.

Method	DSC ↑	HD ↓	HD95 ↓	ASSD ↓	RVE ↓
Ours	0.82±0.06	3.69±0.55	1.83±0.51	0.50±0.23	0.13±0.03

any one task. For quality assessment, decision-layer ordinal calibration improved over raw argmax decoding under the reported validation protocol. For quality enhancement, the three-step UNSB configuration offered better perceptual and distributional scores but lower pixel- and radiomics-based scores than the five-step configuration. For segmentation, EDT supervision and bilateral consistency were integrated with SwinUNETR, ensembling, sagittal test-time augmentation, and light post-processing. These results illustrate how ordinal severity, unpaired domains, and anatomical symmetry can guide method selection in this setting.

The study also has limitations. The available ablations do not provide a complete quantitative comparison with every alternative considered for Task 1a, and the three tasks are evaluated independently on the challenge validation data. Thus, the results should be interpreted as practical challenge evidence rather than as a comprehensive study of representation learning, enhancement objectives, or generalization beyond the LISA data.

References

1. Arnold, T.C., Freeman, C.W., Litt, B., Stein, J.M.: Low-field MRI: Clinical promise and challenges. Journal of Magnetic Resonance Imaging **57**(1), 25–44 (2023)
2. Cardoso, M.J., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
3. Deoni, S.C.L., et al.: Accessible pediatric neuroimaging using a low field strength MRI scanner. NeuroImage **238**, 118273 (2021)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR) (2021)
5. Hao, Y., Liu, M., Yang, H., Li, H., Bai, X., Liao, Y., Qu, H., Tian, Q.: Comprehensive evaluation of unsupervised image enhancement for volumetric fetal brain MRI. In: Medical Imaging with Deep Learning–Short Papers (2025)
6. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: International MICCAI Brainlesion Workshop. pp. 272–284 (2022)
7. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems (NeurIPS) (2017)
8. Hou, L., Yu, C.P., Samaras, D.: Squared earth mover's distance-based loss for training deep neural networks. arXiv preprint arXiv:1611.05916 (2016)
9. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning (ICML). pp. 2127–2136 (2018)

10. Kim, B., Kwon, G., Kim, K., Ye, J.C.: Unpaired image-to-image translation via neural Schrödinger bridge. In: International Conference on Learning Representations (ICLR) (2024)
11. Konz, N., Osuala, R., et al.: Fréchet radiomic distance (FRD): A versatile metric for comparing medical imaging datasets. arXiv preprint arXiv:2412.01496 (2024)
12. Li, Z., Li, Z., Bilgic, B., Lee, H.H., Ying, K., Huang, S.Y., Liao, H., Tian, Q.: DIMOND: DIffusion model OptimizatioN with deep learning. Advanced Science **11**(24), 2307965 (2024)
13. Li, Z., Li, Z., Bilgic, B., Ying, K., Salat, D.H., Polimeni, J.R., Liao, H., Huang, S.Y., Tian, Q.: Noise2average: An iterative residual learning strategy for image denoising without clean data. Imaging Neuroscience **4**, IMAG–a (2026)
14. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988 (2017)
15. Low field pediatric brain magnetic resonance image segmentation and quality assurance: First miccai challenge, lisa 2024. In: Lepore, N., Linguraru, M.G. (eds.) MICCAI Challenge on Low Field Pediatric Brain Magnetic Resonance Image Segmentation and Quality Assurance (LISA 2024). Lecture Notes in Computer Science, vol. 15515. Springer (2025)
16. Low field pediatric brain magnetic resonance image segmentation and quality assurance: Second miccai challenge, lisa 2025. In: Lepore, N., Linguraru, M.G. (eds.) MICCAI Challenge on Low Field Pediatric Brain Magnetic Resonance Image Segmentation and Quality Assurance (LISA 2025). Lecture Notes in Computer Science, vol. 16411. Springer (2026)
17. Liu, M., Liao, Y., Li, H., Zhu, J., Yang, H., Hao, Y., Qu, H., Tian, Q.: Quality-label-free fetal brain MRI quality control based on image orientation recognition uncertainty. Medical Image Analysis p. 103994 (2026)
18. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (ICLR) (2019)
19. Mao, X., Li, Q., Xie, H., Lau, R.Y.K., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 2794–2802 (2017)
20. Marinov, Z., Kleesiek, J., Stiefelhagen, R.: Enforcing anatomical symmetry with euclidean distance transforms for low-field MRI bilateral structure segmentation. In: MICCAI Challenge on Low Field Pediatric Brain Magnetic Resonance Image Segmentation and Quality Assurance (LISA 2025). Lecture Notes in Computer Science, vol. 16411, pp. 63–73. Springer (2026)
21. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing **21**(12), 4695–4708 (2012)
22. Pang, H., Hong, X., Zhang, P., Zhu, P., Yao, S., An, F., Yang, G., Liu, T., Qiu, A., Ye, C., et al.: Cascaded diffusion model and segment anything model for medical image synthesis. Pattern Recognition p. 114148 (2026)
23. Pang, H., Li, X., Yan, T., Ye, C.: Ultra low-field MRI enhancement via conditional diffusion model. In: International Challenge on Ultra-Low-Field MRI Image Enhancement, pp. 9–17. Springer (2025)
24. Pang, H., Xu, S., Zhang, X., An, F., Zhang, X., Wang, Y., Liu, F., Yan, T., Marais, P., Qian, Y., et al.: DRCAM: Dose reduction via contrast amplification modeling for brain contrast-enhanced magnetic resonance images. Biomedical Signal Processing and Control **113**, 109074 (2026)

25. Pang, H., Zhang, P., Hong, X., Chen, S., Ye, C.: D3M: Deformation-driven diffusion model for synthesis of contrast-enhanced MRI with brain tumors. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 151–160. Springer (2025)
26. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European Conference on Computer Vision (ECCV). pp. 319–345 (2020)
27. Shi, X., Cao, W., Raschka, S.: Deep neural networks for rank-consistent ordinal regression based on conditional probabilities. Pattern Analysis and Applications **26**(3), 941–955 (2023)
28. Wang, J., Chan, K.C., Loy, C.C.: Exploring CLIP for assessing the look and feel of images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 2555–2563 (2023)
29. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 586–595 (2018)
30. Zhu, P., Liu, C., Fu, Y., Chen, N., Qiu, A.: Cycle-conditional diffusion model for noise correction of diffusion-weighted images using unpaired data. Medical Image Analysis **103**, 103579 (2025)
31. Zhu, P., Liu, S., Liu, Y., Yap, P.T.: METER: Multi-task efficient transformer for no-reference image quality assessment. Applied Intelligence **53**(24), 29974–29990 (2023)