Controllable Synthesis of Pathological Fetal Brain Ultrasound via Registration-Guided, Radiomics-Driven Latent Diffusion

Richard Zsamboki, Vanda Czipczer, and Alinka Olajos-Horvath

GE HealthCare, Budapest, Hungary
Richard.Zsamboki@gehealthcare.com

Abstract. Fetal brain pathologies, such as hydrocephaly and severe ventriculomegaly, are rare but clinically significant conditions that require accurate prenatal diagnosis. The limited availability of pathological ultrasound images, combined with high variability in imaging due to gestational age, probe angle, and operator technique, poses a major challenge for training robust AI models for diagnostic support tasks. To address this, we propose a controllable synthesis pipeline for transventricular (TV) fetal brain ultrasound that couples registration-guided mask with radiomics-conditioned latent diffusion. First, posterior ventricle and choroid plexus masks from real pathological cases are aligned to normal TV scans using a skull-anchored affine-then-diffeomorphic registration strategy, ensuring plane-specific anatomical localization. Next, we apply radiomics-conditioned inpainting in a VQ-VAE latent space using a Latent Brownian Bridge Diffusion Model (LBBDM). The model is conditioned on (i) the masked normal image, (ii) fixed intensity prior for key structures, and (iii) radiomics texture features sampled from the training distribution, yielding realistic and controllable pathology appearance. In a downstream evaluation for hydrocephaly image similarity-based retrieval, augmenting real-world training data with our synthetic images increases the expected number of diagnosis-matching references in a top 10 query from 0.35 to 0.7 with a self-supervised ResNet-18 encoder. This work demonstrates that incorporating synthetic data boosts retrieval performance for rare fetal pathologies.

Keywords: Fetal ultrasound · Image synthesis · Image retrieval.

1 INTRODUCTION

Hydrocephaly and severe ventriculomegaly are relatively infrequent, yet timely identification is critical for counseling and care planning [1]. Prenatal assessment of the fetal brain relies on standardized sonographic planes, among which the transventricular (TV) view is key for evaluating posterior ventricles (PV) and choroid plexus (CP)[2]. In practice, three factors hinder learning-based decision support: (1) scarcity of pathological TV images, (2) substantial appearance variability due to gestational age, fetal pose, and acquisition settings, and (3) incomplete annotations and ontology coverage in real-world datasets [2–5].

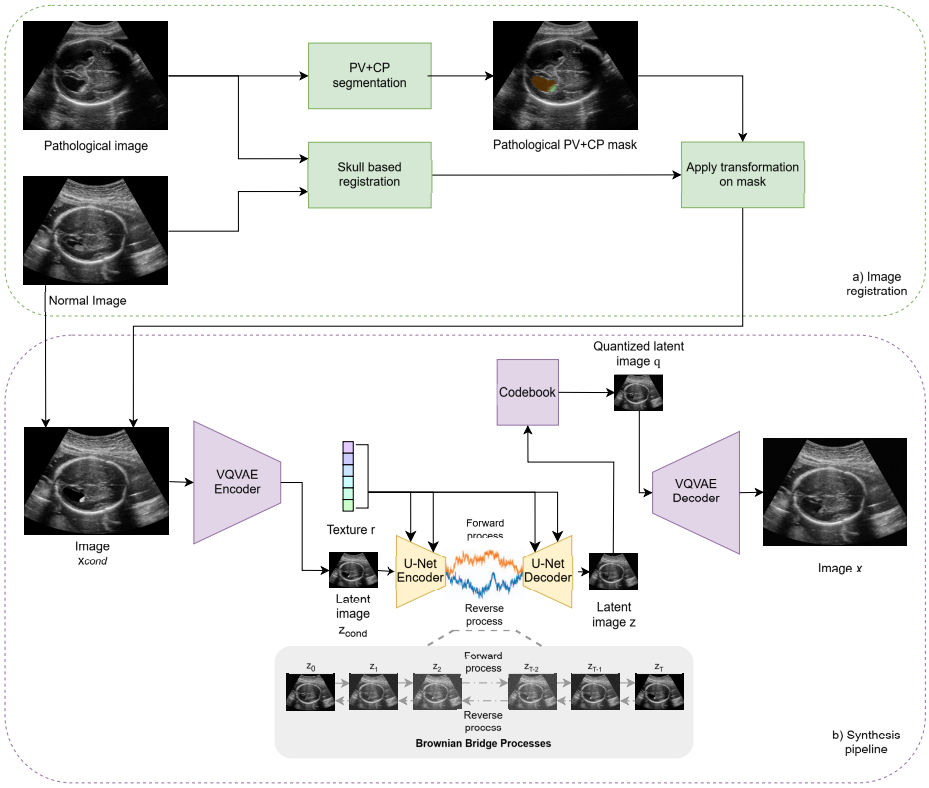

Fig. 1. Overview of the proposed synthesis pipeline. a) demonstrates that the pathological and its PV+CP mask are registered to a normal image via a skull-based affine-then-diffeomorphic image registration, then in b) using the registered mask, the normal image is masked to form x_{cond} and encoded by a VQ-VAE into the conditional latent z_{cond}. From z_{cond}, a brownian-bridge process guided by r synthesizes pathology only within the masked region. The synthesized latent is decoded to image space to yield anatomically anchored, texture-controllable pathological exemplars.

Synthetic data is a promising strategy for countering class imbalance in medical imaging [3]. Recent diffusion-based ultrasound models explore controllable medical image synthesis using DDPMs guided by semantic labels, such as [6, 7], demonstrating the feasibility of generating useful images for downstream segmentation. In the fetal domain, Duan et al. [3] introduced FetalFlex, an anatomy-guided diffusion model enabling controllable multi-plane fetal ultrasound synthesis. Parallel advances have focused on CT/MRI. Kim et al. [8] proposed a radiomics conditioned tumor generator that combines a GAN-based shape generator with a Brownian Bridge Diffusion Model - first introduced in [9] - in a VQVAE [11] latent space, enabling tumor generation based on user-specified radiomics features such as size, shape and texture at an arbitrary location. Their

results highlight the strength of radiomics features as semantically meaningful controls for generative models.

However, existing fetal ultrasound synthesis models generally lack localized, anatomy-aware pathology synthesis, and radiomics-controlled diffusion has not yet been translated to fetal brain ultrasound. Unlike CT/MRI tumor synthesis, PV/CP pathology must be inserted into a standardized TV plane while remaining anatomically consistent with the skull outline, midline, and surrounding brain structures. To address this, we introduce a controllable synthesis pipeline for pathological TV ultrasound images of hydrocephaly and ventriculomegaly (Fig. 1). The method transplants real PV/CP masks from pathological cases into normal TV scans using skull-anchored affine-diffeomorphic registration with Hausdorff-based quality assurance, computes higher-order radiomics texture features [15], and uses a dual-conditioned LBBDM [8] to inpaint pathology within the transplanted anatomy. Finally, we show that the resulting synthetic images improve hydrocephaly and ventriculomegaly similarity retrieval.

2 Methods

2.1 Registration-Guided Mask Transplant

Inputs comprise (1) a normal TV ultrasound image x_n and a pathological image x_{pat}, (2) source pathological masks $\mathcal{M}_{\text{pat}} = \{M_{\text{PV}}, M_{\text{CP}}\}$ delineated on pathological TV images and (3) a head mask (skull outline and midline/falx) for x_n and x_{pat}. We treat x_n as fixed and x_{pat} as moving. Before registration, all mask-transplant images were assigned manual orientation labels to align fixed and moving TV images consistently. Ambiguous cases were excluded.

Let d_H denote the skull-outline Hausdorff distance, with stage-wise values d_{H1} and d_{H2}. Thresholds are $\tau_1 = 5.0$ pixels and $\tau_2 = 4.0$ pixels.

We align x_{pat} onto x_n through three stages (Fig. 2). First, a skull mask-guided affine stage computes two heuristic initializations of in-plane rotation, translation, and scale, optimizes both, and keeps the lower-d_{H1} solution. Pairs with $d_{H1} > \tau_1$ are rejected. Second, an intensity-refined affine step maximizes local similarity (e.g., NCC) from the accepted T_{aff}, but keeps T_{int} only if d_{H2} remains below τ_1 and does not worsen relative to d_{H1}. Third, if $d_{H2} \leq \tau_2$, mask-driven diffeomorphic registration [10] captures residual non-rigid differences. Otherwise the accepted affine transform is retained. The final transform T^* maps $\mathcal{M}_{\text{pat}}$ into the x_n frame.

2.2 Preprocessing

For synthesis, images are resampled to 512×384, and the intensities are scaled by mapping the 0–99.5th percentile range to $[-1, 1]$. Additionally, we assign fixed intensity priors to PV/CP regions in the preprocessed image x (e.g., $+1$ for CP, -1 for PV) to mark the anatomical region to be reconstructed by VQ-VAE and synthesized by LBBDM, producing the conditional image x_{cond}. Furthermore,

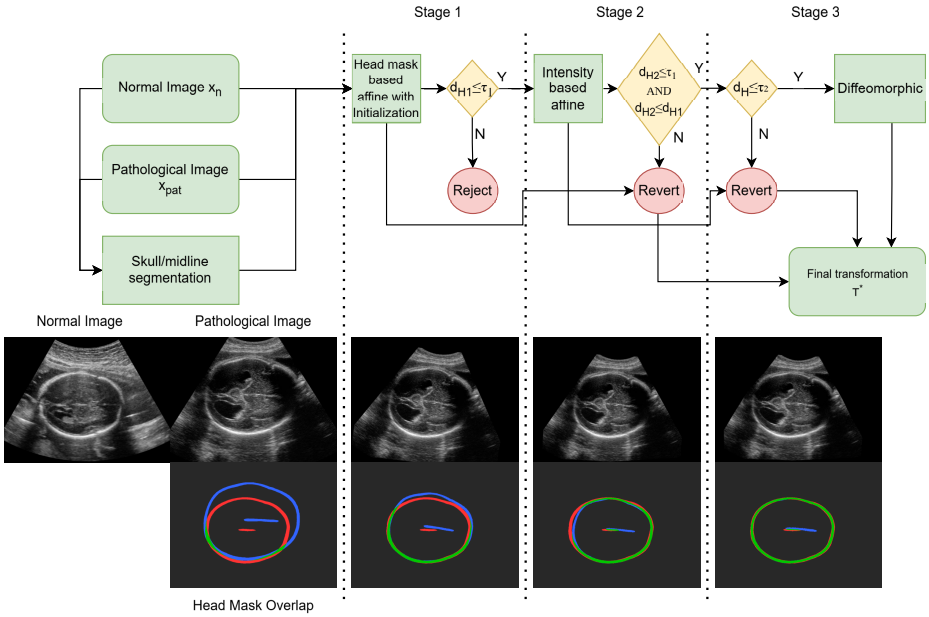

Fig. 2. Mask–based registration with quality control and fallbacks. Top: registration workflow, Bottom: example fixed (normal) and moving (pathological) image pair. For each stage, the top row shows the transformed moving image, and the bottom row shows the head-mask overlay (blue: transformed moving mask, red: fixed mask, green: overlap).

we compute higher-order texture features to create radiomics vectors r from the combined PV/CP region using automatic segmentations from the TV training set, excluding manually annotated pathological source masks. At synthesis time, r is sampled from this empirical training distribution for conditioning.

For the downstream task, the input images are resampled to 256x256, continuing with their intensities initially normalized to $[0, 1]$, then with predefined mean value of 0.183 and standard deviation of 0.173.

2.3 VQ-VAE

We employ a conditional, 2D VQ-VAE [12] to encode x_{cond} into a latent representation and to decode synthesized latent back to image space. The encoder $E(\cdot)$ applies two convolutional downsampling stages (feature widths 128 and 256) to map the normalized image x_{cond} to a quantized latent grid (codebook size 8192, dimension 128). The decoder $D(\cdot)$ mirrors this architecture to reconstruct x from the quantized latent. Training minimizes a mask-weighted reconstruction

objective [13, 14]:

$$\mathcal{L}_{\text{rec}} = \lambda_1 \cdot \mathcal{L}_1(x, x_{\text{cond}}) + \lambda_2 \cdot \mathcal{L}_{\text{codebook}} + \lambda_3 \cdot (1 - \text{SSIM}(x, x_{\text{cond}})) + \lambda_4 \cdot \text{LPIPS}(x, x_{\text{cond}}) \tag{1}$$

Here, SSIM captures local structural consistency, LPIPS provides a VGG-based perceptual distance [14], and the PV/CP union receives increased weight to emphasize reconstruction of the clinically relevant region.

We trained the 2D VQ-VAE with an Adam optimizer (learning rate 1×10^{-4}), using a mini batch size of 10 and gradient accumulation over 6 iterations (effective batch size 60). The model was optimized for 175k steps, and we selected the checkpoint with the lowest validation loss, reached at 126k steps. The reconstruction loss weights were set empirically to ($\lambda_1 = 1.0$), ($\lambda_2 = 5.0$), ($\lambda_3 = 0.1$), and ($\lambda_4 = 0.1$), with a ($\times 10$) weighting applied inside the union of posterior ventricle and choroid plexus masks.

2.4 Inpainting with LBBDM

We model anatomy- and texture-controlled synthesis in the latent space of a pretrained VQ-VAE using a latent Brownian Bridge diffusion model (LBBDM). The frozen encoder maps the original image and its mask-conditioned counterpart to latent codes ($z = E(x)$) and ($z_{\text{cond}} = E(x_{\text{cond}})$), which we standardize channel-wise using dataset-wide means and standard deviations. LBBDM is then trained entirely in this latent space, keeping the VQ-VAE fixed.

Unlike DDPMs, LBBDM learns a Brownian-bridge trajectory that directly connects the conditional latent (z_{cond}) to the target latent (z), with bounded variance that keeps samples close to the conditional anatomy. A 2D UNet operates on the latent grid and predicts the analytic bridge direction from the noisy latent state, time point, conditional latent, and radiomics vector using an $L1$ loss.

At inference, reverse-time updates start from z_{cond} and produce a target-like latent sample, which the VQ-VAE decoder maps to a synthetic pathological image respecting the anatomical mask prior and radiomics-derived texture profile.

The LBBDM training was done for 300k steps with a batch size of 16, with initial learning rate of 5×10^{-6}, using Adam optimizer and applied a ReduceLROnPlateau scheduler (factor 0.5, patience 3000) and setting the minimum learning rate of 5×10^{-7}. The Brownian bridge objective was optimized over 1000 diffusion timesteps, and an exponential moving average of the network weights (decay 0.995) and 200 timesteps were used for sampling.

2.5 Downstream task

For image similarity retrieval, we pretrained a ResNet-18 [16] encoder in a self-supervised contrastive manner using a SimCLR-style NT-Xent loss (temperature of 0.1) [17] with ultrasound appropriate augmentations (e.g. random crop, flip, rotation, contrast change, Gaussian noise and blur) and a small MLP projection head.

We compare two training regimes: (1) real-world data (RWD) only, and (2) a mixture of RWD and synthetic images. In the mixed regime, each batch of size 64 contained equal numbers of real and synthetic images. We optimized with AdamW (learning rate 3×10^{-3}, weight decay 10^{-4}) and applied a ReduceLROnPlateau scheduler (factor 0.5, patience 5), evaluating the contrastive loss on the held-out RWD validation set every 200 training steps and pretraining was run for a total of 10k steps.

3 Experimental results

3.1 Datasets and Annotations

We curate ∼135k unlabeled ultrasound images from the SUOG project spanning multiple prenatal views. Using feature clustering and off-the-shelf in-house view classifiers, we extracted 724 TV images for training and validation. Random manual spot-checks confirmed reliable TV-plane selection, although exhaustive review was not performed. From a labeled cohort (∼8.6k images, ∼490 patients, 113 diagnoses), ventriculomegaly/hydrocephaly cases were curated to 67 good-quality fetal brain images with visible pathology.

Off-the-shelf inhouse segmentation models automatically contoured PV, CP and head masks on training and validation data, and head masks for ventriculomegaly and hydrocephaly data. For registration-guided mask transplant, PV and CP were manually delineated on 39 pathological TV images (10 hydrocephaly, 29 ventriculomegaly) where both structures were identifiable, using a SAM2-assisted [18] semi-automatic tool. These annotations served only as source masks for spatial placement. Radiomics vectors were derived from the broader TV training set using automatic PV/CP segmentations.

The TV dataset was split 90/10 for VQ-VAE, LBBDM, and ResNet-18 training/validation. Synthetic images were generated only from the training split and used only for mixed SSL pretraining, never for validation or retrieval evaluation. Downstream augmentation used ∼6k hydrocephaly and ∼16k ventriculomegaly synthetic images, with equal real/synthetic samples per mixed SSL batch and patient-disjoint real-world retrieval evaluation (Section 3.3).

3.2 Implementation details

All models were implemented in Python using PyTorch and MONAI. The training of VQ-VAE and LBBDM was executed on a single NVIDIA A100 40 GB GPU, while the SSL pretraining of ResNet-18 was done on a single NVIDIA RTX 5000 16GB GPU. Experiments were run on a Linux workstation, and all hyperparameters (learning rates, schedulers, and training steps) are reported in the corresponding method descriptions.

3.3 Evaluation Protocol

We evaluate image similarity retrieval using the encoder from Section 2.5. Feature embeddings are extracted for the labeled cohort (containing ∼8.6k images,

∼490 patients, 113 diagnoses) from Section 3.1 and for a patient disjoint query set (i.e., no patient overlap between reference and query). For each query image, cosine similarity is computed against all reference embeddings, and the top-k nearest neighbors are retrieved (k=10 by default).

Retrieval quality is assessed against ground-truth diagnosis labels using Average Diagnosis Match Success (ADMS), concept first introduced in [19], defined as the mean number of top-10 retrieved images that share the query's diagnosis (range 0–10). We report ADMS on a patient-disjoint query set, separately for hydrocephaly and ventriculomegaly, and compare the training regimes in Section 2.5.

3.4 Quantitative Results

Table 1 shows that SSL pretraining with both RWD and synthetic data improves similarity retrieval compared to RWD only: ADMS increased from 0.35 to 0.70 for hydrocephaly and from 1.92 to 2.22 for ventriculomegaly. The 95% bootstrap CIs shift upward for both diagnoses, while the slightly wider intervals reflect higher case-level variability.

Table 1. ADMS summary statistics for hydrocephaly and ventriculomegaly. n denotes real query images per diagnosis. SD denotes standard deviation. 95% CI is reported as lower and upper bounds using bootstrap resampling (10,000 resamples, seed=42).

Training data	Hydrocephaly			Ventriculomegaly		
	n	Mean±SD	95% CI	n	Mean±SD	95% CI
RWD	37	0.35 ± 0.63	[0.16, 0.57]	63	1.92 ± 2.17	[1.4, 2.48]
RWD + Synthetic	37	**0.70** ± 1.00	[0.41, 1.03]	63	**2.22** ± 2.23	[1.68, 2.78]

3.5 Qualitative Results

Fig. 3 shows typical inpainting results for ventriculomegaly and hydrocephaly in the TV plane. The method preserves plane-consistent anatomy (e.g. midline, cavum septum pellucidum) and reproduces key signs, including ventricle dilatation and the choroid plexus "dangling" sign, with smooth echotexture transitions at mask boundaries.

By design, the method modifies only the masked PV/CP region while preserving surrounding anatomy. This scoped inpainting targets localized TV-plane pathological exemplars for retrieval augmentation rather than full biomechanical simulation of severe hydrocephaly. Therefore, global morphological changes such as cortical mantle thinning or midline shift are not reproduced. Despite this constraint, the synthesized images capture diagnostically relevant local signs that drive TV-plane similarity retrieval.

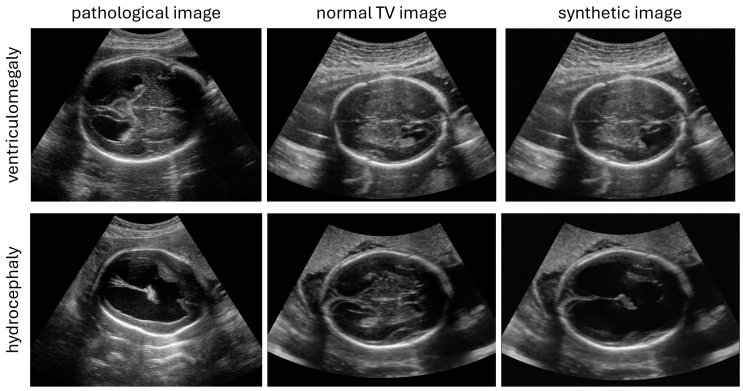

Fig. 3. Synthesis in the TV plane. For each pathology: (left) pathological image, (middle) normal TV image, (right) generated synthetic image. The method reproduces the posterior ventricle dilation and dangling choroid plexus.

4 Conclusion

We presented a controllable synthesis framework for pathological fetal brain ultrasound in the transventricular plane that couples skull-anchored registration of real PV/CP masks with radiomics-conditioned latent Brownian-bridge diffusion. The method generates localized hydrocephaly and ventriculomegaly appearances while preserving surrounding TV-plane anatomy, and improves patient-disjoint similarity retrieval when added to training data (ADMS increased from 0.35 and 1.92 to 0.70 and 2.22, respectively). The main limitations are the modest pathological cohort, and the localized nature of the current inpainting strategy, which does not model global morphological effects of severe hydrocephaly. Future work will address these limitations through reader studies to assess clinical utility, quantitative assessment of radiomics controllability and clinical realism, and synthesis models that couple local PV/CP inpainting with global TV-plane anatomical deformation.

5 Acknowledgements

Data collection and annotation activities were part of the SUOG Project (www.suog.org), an EIT Health Innovation supported project.

References

1. Norton, M.E., Fox, N.S., Monteagudo, A., Kuller, J.A., Craigo, S.: Fetal ventriculomegaly. American Journal of Obstetrics and Gynecology 223(6), B30–B33 (2020)
2. Malinger, G., Paladini, D., Haratz, K., Monteagudo, A., Pilu, G.L., Timor-Tritsch, I.E.: ISUOG Practice Guidelines (updated): sonographic examination of the fetal

central nervous system. Part 1: performance of screening examination and indications for targeted neurosonography. Ultrasound in Obstetrics & Gynecology 56(3), 476–484 (2020)
3. Duan, Y., Tan, T., Zhu, Z., Huang, Y., Zhang, Y., Gao, R., Pang, P.C.-I., Gao, X., Tao, G., Cong, X., Li, Z., Liang, L., He, G., Yin, L., Deng, X., Yang, X., Ni, D.: FetalFlex: Anatomy-Guided Diffusion Model for Flexible Control on Fetal Ultrasound Image Synthesis. arXiv:2503.14906 (2025).
4. De Robertis, V., Bosco, M., Volpe, P.: Ultrasound Evaluation of the Fetal Brain. In: *The Global Library of Women's Medicine (GLOWM), Ultrasound in Obstetrics* (2024). https://www.glowm.com/article/heading/vol-18–ultrasound-in-obstetrics–ultrasound-evaluation-of-the-fetal-brain/id/419303
5. Meyer, A., Murali, A., Zarin, F., Mutter, D., Padoy, N.: UltraSam: A foundation model for ultrasound using large open-access segmentation datasets. *International Journal of Computer Assisted Radiology and Surgery* (2025). https://doi.org/10.1007/s11548-025-03517-8
6. Stojanovski, D., Hermida, U., Lamata, P., Beqiri, A., Gomez, A.: Echo from noise: synthetic ultrasound image generation using diffusion models for real image segmentation. In: International Workshop on Advances in Simplifying Medical Ultrasound, pp. 34–43. Springer (2023)
7. Czipczer, V., Timkó, E., Zsámboki, R., Takács, P., Kiss, E., Prosszer, M., Várkonyi, D.: Power in small hybrid dataset in training deep-learning models: how to focus annotation effort. In Medical Imaging 2025: Imaging Informatics vol. 13411, pp. 89-93. SPIE (2025)
8. Kim, J., Na, I., Ko, E.S., Park, H.: Tumor synthesis conditioned on radiomics. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 3635–3646. IEEE (2025)
9. Li, B., Xue, K., Liu, B.: BBDM: Image-to-Image Translation with Brownian Bridge Diffusion Models In Proceedings of the IEEE/CVF conference on computer vision and pattern Recognition, pp. 1952–1961. IEEE (2023)
10. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric diffeomorphic image registration with the demons algorithm. In International conference on medical image computing and computer-assisted intervention, pp. 319–326, Springer (2007)
11. Van Den Oord, A., Vinyals, O., Kavukcuoglu, K.: Neural discrete representation learning In Advances in neural information processing systems, NIPS (2017)
12. Peng, J., Liu, D., Xu, S., Li, H.: Generating diverse structure for image inpainting with hierarchical VQ-VAE. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10775–10784. IEEE (2021)
13. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* **13**(4), 600–612 (2004)
14. 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586–595 (2018)
15. Van Griethuysen, J.J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G., Fillion-Robin, J.C., Pieper, S. and Aerts, H.J.: Computational radiomics system to decode the radiographic phenotype. In Cancer Research, 77(21), pp.e104–e107. AACR (2017)

16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. IEEE (2016)
17. 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)
18. Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Wang, H., Feichtenhofer, C., Sauer, A., Kirillov, A.: SAM 2: Segment Anything in Images and Videos. arXiv preprint arXiv:2408.00714 (2024)
19. Takacs, P., Zsamboki, R., Kiss, E., Dhombres, F.: AI Models for Ultrasound Image Similarity Search: A Performance Evaluation. In BIOSTEC (2): HEALTHINF, pp. 414–424. (2025)
20. Wang, T., Isola, P.: Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In: International Conference on Machine Learning, pp. 9929–9939. PMLR (2020)