
SFL-NET: SOURCE-FACTORIZED LATENT REPRESENTATION LEARNING FOR MULTI-CONTRAST MRI TO TAU-PET SYNTHESIS

Agamdeep S. Chopra^{1,*}, Caitlin Neher^{1,†}, Tianyi Ren^{1,†}, Juampablo E. Heras Rivera¹,
Hesamoddin Jahani², and Mehmet Kurt¹

¹Department of Mechanical Engineering, University of Washington, Seattle, WA, USA

²Department of Radiology, University of Washington, Seattle, WA, USA

[†]These authors contributed equally

*Correspondence: achopra4@uw.edu

{achopra4,neherc,tr1,jehr,hesamj,mkurt}@uw.edu

July 7, 2026

ABSTRACT

Tau positron emission tomography supports Alzheimer’s disease staging but is difficult to scale because of tracer, scanner, and radiation constraints. Synthesis from structural MRI is therefore attractive, but it is a particularly difficult setting. T1-weighted and FLAIR MRI provide anatomy and disease correlated morphology, but they do not directly measure Tau-PET relevant signal. We introduce SFL-Net, a multi-input synthesis framework that predicts Tau-PET from T1-weighted and FLAIR MRI. SFL-Net factorizes the latent representation into shared, T1-specific, FLAIR-specific, and complementary pathways and preserves anatomical detail through latent structural conditioning rather than direct encoder-decoder connections. We evaluated SFL-Net and baseline models using 605 training and 83 validation subjects from ADNI-3 and OASIS-3 datasets. Evaluation included raw image fidelity, standardized uptake value ratio agreement, high uptake overlap, regional Bland-Altman bias, braak derived stage agreement, non-inferiority sensitivity analysis, and latent component Shapley attribution.

SFL-Net performed competitively on both clinically relevant and reconstruction metrics, while also delivering explicit source level auditability that conventional UNet derived models lack.

Keywords Tau-PET, MRI, vector quantization, source-factorized representation learning, interpretable deep learning, Alzheimer’s disease, medical image synthesis.

1 Introduction

Tau positron emission tomography (tau-PET) provides in-vivo information about neurofibrillary pathology and is closely related to Alzheimer’s disease (AD) stage and clinical decline [1–3]. However, tau-PET is difficult to scale because it requires radioactive tracer injection and exposure, long acquisition, and specialized imaging infrastructure. Structural MRI is more widely available and provides high resolution anatomical information, motivating MRI to PET synthesis as an attractive alternative [4–6].

Prior tau-PET synthesis studies have considered several cross-modal input settings, but these settings differ substantially in clinical burden and information content. For ex-

ample, tau-PET synthesis from FDG-PET or amyloid-PET can be more accurate than synthesis from structural MRI because the input already contains molecular or metabolic information related to AD pathophysiology [6]. However, PET to PET synthesis only partially addresses the scalability problem, since FDG-PET and amyloid-PET still require radioactive tracers, PET acquisition, radiation exposure, and specialized infrastructure. An alternative approach is to synthesize tau-PET from structural MRI, requiring the model to infer molecular tau load based on anatomy and disease associated morphological changes rather than using another PET-based biomarker. This constitutes a more weakly constrained task. T1-weighted MRI primarily reflects brain structure and atrophy, and FLAIR highlights

tissue abnormalities such as white matter hyperintensities, vascular damage, and other structural signal alterations, but neither imaging sequence provides a direct measure of tau binding. Existing structural MRI to tau-PET work has demonstrated feasibility, but has largely focused on T1-weighted MRI alone or on conventional CNN and UNet based synthesis models [6, 7]. Consequently, it remains unclear how well an interpretable structural MRI to tau-PET model performs when evaluated against stronger capacity-matched synthesis baselines and clinically oriented tau metrics.

Medical image synthesis is usually evaluated with voxel-wise error or perceptual similarity, but these metrics are insufficient for tau-PET. Quantitative utility depends on regional standardized uptake value ratio (SUVR), preservation of sparse high-uptake signal, calibration across disease stages, and downstream agreement with tau staging rules. A visually plausible synthetic PET image can still fail clinically if regional uptake is biased toward the cohort mean or if advanced stage subjects do not cross SUVR thresholds. These issues are especially relevant for AD, where molecular burden is spatially heterogeneous and clinically interpreted through regional staging rather than global image similarity alone.

A second limitation is interpretability. UNet style synthesis models can obtain strong average metrics but do not distinguish information shared across MRI contrasts from input-specific or interaction dependent cues. Direct encoder-decoder skip connections can also transmit high bandwidth anatomical information around the bottleneck, making it difficult to attribute the generated PET signal to specific latent mechanisms. This is problematic in multi-input synthesis, where the central question is not only whether the target can be predicted, but which source information supports the prediction. This concern is especially relevant because UNet-style skip connections were introduced to recover spatial detail [8], and image to image translation work has used them specifically to pass low level structure around the bottleneck [9].

Thus, we propose SFL-Net, a Source-Factorized Latent Network for interpretable tau-PET synthesis from T1-weighted and FLAIR MRI. SFL-Net addresses the limitations above by replacing a monolithic continuous bottleneck with a quantized latent representation organized into shared, T1-specific, FLAIR-specific, and cross-source complementary pathways. The method has two core components. First, a vector-quantized encoder is regularized with a source-factorization loss that shapes the dependence structure among latent partitions [10]. Second, a structure-conditioned latent decoder maintains anatomical fidelity without relying on any direct feature skip connections between encoder and decoder. This design aligns with conditioning strategies that introduce information via controlled feature modulation or spatial structural priors [11–14]. The purpose of the design is to enable operational, source-level auditability via component ablation and attribution that can be integrated with state-of-the-art synthesis frameworks.

Our contributions are: (i) a source-factorized quantized latent architecture for structural MRI to PET synthesis; (ii) a structure-conditioned latent decoder that preserves anatomical detail without direct encoder-decoder skip connections; and (iii) a clinically oriented validation that includes PET fidelity, SUVR agreement, high-uptake overlap, braak-derived stage agreement, Bland-Altman analysis, non-inferiority sensitivity analysis, and latent-component Shapley attribution against strong baselines.

2 Method

2.1 SFL-Net Architecture

SFL-Net is composed of a shared/source-specific encoder for the separate T1 and FLAIR branches, a cross-source complementary encoder that processes concatenated T1/FLAIR inputs, and a structure-conditioned latent decoder (Fig. 1). Each encoder includes a quantization layer that performs nearest neighbor code assignment, mapping the encoder’s latent output to a code vector from its learned vector codebook [Fig. 1] [10]. The shared/source-specific encoder processes T1 and FLAIR with shared weights and produces indexed latent partitions. The cross-source complementary encoder receives the concatenated inputs and captures interaction dependent information that cannot be derived from either single-source pathway. The decoder receives quantized latent factors as its primary input and reconstructs the PET volume while being conditioned by edge-derived structural cues. This design avoids full-resolution encoder feature skip connections so that PET-relevant signal is forced to pass primarily through the factorized latent representation. The latent partitions are:

$$z_{T1} \rightarrow [z_{T1}^s, z_{T1}^u], \quad z_{FL} \rightarrow [z_{FL}^s, z_{FL}^u], \quad z_C \rightarrow z^c, \quad (1)$$

where z_{T1}^s, z_{FL}^s are shared factors between T1 and FLAIR input sources containing mutual subject information such as anatomy and structure, z_{T1}^u, z_{FL}^u are source-specific factors that contain orthogonal information between the input sources, and z^c is the cross-source complementary factor containing information that only arises by coupling the input sources.

2.2 Latent Source Factorization

The architectural streams encourage separation of shared, source-specific, and cross-source complementary information in the latent representation. The shared, source-specific, and complementary terminology is conceptually related to partial information decomposition, but SFL-Net does not estimate formal PID atoms; the factorization is an operational representation constraint designed for multi-input synthesis, ablation, and component-level attribution [15, 16]. Shared factors are encouraged to agree across inputs, source-specific factors are discouraged from sharing information, and the cross-source complementary factor is discouraged from collapsing into either shared or source-specific streams.

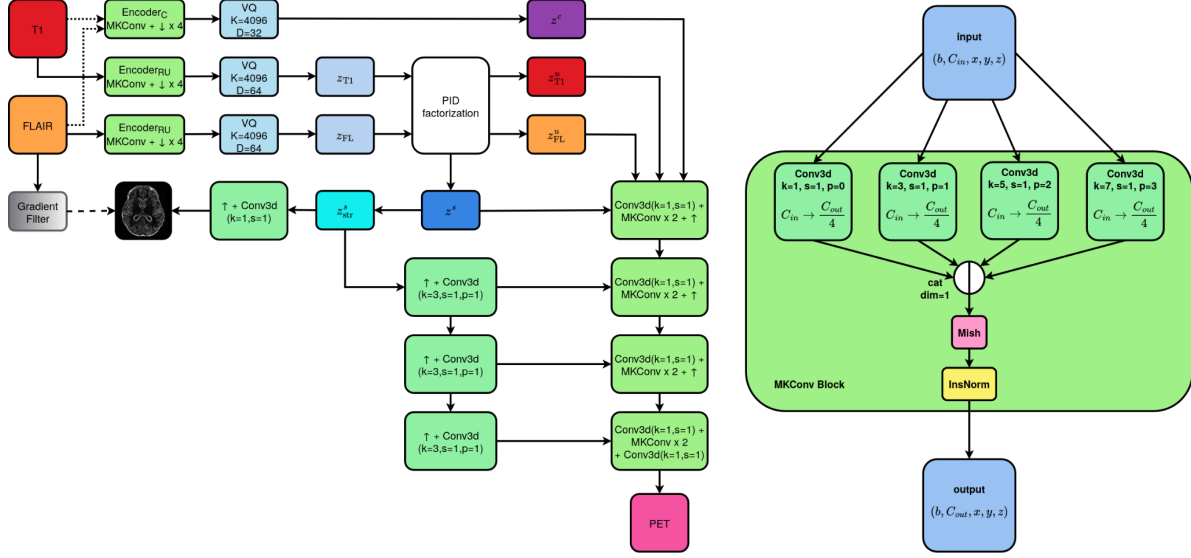


Figure 1: General architecture of the proposed SFL-Net model for structural MRI to tau-PET synthesis. T1-weighted and FLAIR MRI are encoded into shared, source-specific, and cross-source complementary quantized latent pathways. Anatomical structure is provided through low-bandwidth latent-derived conditioning rather than direct encoder-decoder feature concatenation. $\uparrow$ refers to trilinear interpolation of desired scale, and $\downarrow$ refers to convolution downsample of kernel size 2 and stride 2.

For each pair of latent partitions, we estimate discrete mutual information from differentiable soft assignments to the corresponding codebooks [17–19] and normalize it by the finite-alphabet upper bound. For pre-quantized latent z_P and codebook $\mathcal{E}_P = \{e_{P,k}\}_{k=1}^{K_P}$,

$$\pi_P(k | z_P) = \text{softmax}_k \left(-\frac{\|z_P - e_{P,k}\|_2^2}{\tau} \right). \quad (2)$$

Given paired samples from partitions A and B , their joint code-index distribution is approximated as

$$\hat{p}_{AB}(k, \ell) = \frac{1}{N} \sum_{n=1}^N \pi_A(k | z_A^{(n)}) \pi_B(\ell | z_B^{(n)}), \quad (3)$$

with marginal code-index distributions obtained by summing over the joint distribution,

$$\hat{p}_A(k) = \sum_{\ell=1}^{K_B} \hat{p}_{AB}(k, \ell), \quad \hat{p}_B(\ell) = \sum_{k=1}^{K_A} \hat{p}_{AB}(k, \ell). \quad (4)$$

We then compute the discrete mutual information between the induced code-index variables [17],

$$I(A; B) = \sum_{k=1}^{K_A} \sum_{\ell=1}^{K_B} \hat{p}_{AB}(k, \ell) \log \frac{\hat{p}_{AB}(k, \ell) + \epsilon}{(\hat{p}_A(k) + \epsilon)(\hat{p}_B(\ell) + \epsilon)}. \quad (5)$$

We can thus define a bounded finite-alphabet normalized MI score, capacity normalized mutual information (cNMI) as

$$\text{cNMI}(A; B) = \frac{I(A; B)}{\log(\min(K_A, K_B)) + \epsilon}. \quad (6)$$

that follows from the finite-alphabet bound $I(A; B) \leq \min\{H(A), H(B)\} \leq \log(\min(K_A, K_B))$ [17], yielding $\text{cNMI}(A; B) \in [0, 1]$. Unlike entropy based NMI [20, 21], the denominator is fixed by the codebook capacity, giving a stable regularization scale across latent partition pairs, including cases where $K_A \neq K_B$ [22].

Finally, we define the source-factorization loss as

$$\begin{aligned} \mathcal{L}_{\text{SF}} = & \lambda_u \text{cNMI}(z_{\text{T1}}^u; z_{\text{FL}}^u) + \lambda_s \mathcal{R}_s + \lambda_{us} \sum_{i,j} \text{cNMI}(z_i^u; z_j^s) \\ & + \lambda_{cs} \sum_j \text{cNMI}(z^c; z_j^s) + \lambda_{cu} \sum_i \text{cNMI}(z^c; z_i^u), \end{aligned} \quad (7)$$

$$\mathcal{R}_s = 1 - \text{cNMI}(z_{\text{T1}}^s; z_{\text{FL}}^s) + \lambda_{\text{align}} \|z_{\text{T1}}^s - z_{\text{FL}}^s\|_2^2. \quad (8)$$

Here, A and B denote two distinct latent partitions; P is the partition index such that $z_P \in \{z_{\text{T1}}^s, z_{\text{T1}}^u, z_{\text{FL}}^s, z_{\text{FL}}^u, z^c\}$; $\mathcal{E}_P = \{e_{P,k}\}_{k=1}^{K_P}$ is its codebook, where K_P is the number of code entries and K_A, K_B are the codebook sizes for the compared partitions; $\pi_P(k | z_P)$ is the soft assignment of z_P to code k ; τ controls assignment sharpness; M is the number of paired latent samples; $\hat{p}_{AB}$ and $\hat{p}_A, \hat{p}_B$ are the estimated joint and marginal code-index distributions; $I(A; B)$ is the corresponding discrete mutual information; and ϵ is used for numerical stability. i and $j \in \{\text{T1}, \text{FL}\}$. The shared alignment term, $\|z_{\text{T1}}^s - z_{\text{FL}}^s\|_2^2$, can be omitted by setting $\lambda_{\text{align}} = 0$ when agreement is defined only in terms of code-index dependence. In this work, we set $\lambda_{\text{align}} = 1$ to encourage both shared code-index information and latent content agreement. To reduce latent under utilization chan-

nelwise collapse in the source-specific and complementary factors, we applied a variance-floor penalty to z_{T1}^u , z_{FL}^u , and z^c :

$$\mathcal{L}_{vf}(z) = \frac{1}{C} \sum_{j=1}^C \max(0, \sigma_0 - \text{Std}(z_j)), \quad (9)$$

where σ_0 is a predefined minimum standard deviation threshold, C denotes the number of channels, and z represents the latent factor. This penalty discourages low variance latent channels and helps preserve effective capacity in non-shared partitions. To reduce asymmetry between the shared branches, we used random input swapping and stochastic shared factor mixing during training to obtain the shared latent partition as

$$z^s = \rho z_{T1}^s + (1 - \rho) z_{FL}^s, \quad \rho \sim \mathcal{U}(0, 1). \quad (10)$$

These operations were used only as training-time regularizer and were disabled at inference and during inference $z^s = z_I^s$, where I is the first source latent by input channel index.

2.3 Structure-Conditioned Latent Decoding

UNet derived skip connections can preserve spatial detail but also create high bandwidth pathways around the bottleneck, which can weaken attribution to the latent vectors [8]. To counter this, SFL-Net conditions the decoder using low bandwidth structural maps predicted from a fixed channel slice of the shared latent representation. This slice is conditioned using an input derived edge map during training supervision. We thus reserve a fixed set of channels $\mathcal{I}_{str}$ within z^s for structural conditioning and define

$$z_{str} = z_{str}^s[:, \mathcal{I}_{str}, :], \quad (11)$$

where the channel indices and dimensionality of z_{str}^s is fixed across training and inference. The decoder is initialized at the coarsest scale by

$$d_L = \phi_L(\eta(z^s, z_{T1}^u, z_{FL}^u, z^c)), \quad (12)$$

where $\eta(\cdot)$ denotes latent fusion. At decoder level ℓ , a scale-matched structural conditioning map is predicted from z_{str} and fused with the upsampled decoder feature:

$$\tilde{q}_\ell = \psi_\ell(z_{str}^s), \quad d_\ell = \phi_\ell(\text{Up}(d_{\ell+1}) \parallel \tilde{q}_\ell), \quad (13)$$

where $\parallel$ indicates concatenation along the channel dimension, ϕ_ℓ denotes the decoder block, and ψ_ℓ corresponds to the sequence of layers employed to produce the structural map at the associated decoder level.

For supervision, a fixed 3D Sobel edge magnitude map is computed from the structural MRI input and min-max normalized to obtain $\tilde{M}$ [23, 24]. Structural consistency is enforced in z_{str} by

$$\mathcal{L}_{edge} = \left\| \tilde{M} - \hat{M} \right\|_2^2, \quad \hat{M} = \theta(z_{str}^s). \quad (14)$$

Where θ is a shallow projection layer only activated during training (Fig. 1). Thus, anatomical structure regularizes the latent-derived conditioning pathway without directly providing edge maps or encoder feature maps to the decoder.

2.4 Total Training Objective

The full SFL-Net objective combines PET reconstruction, vector-quantization regularization, latent source-factorization, and structural consistency losses:

$$\mathcal{L} = \lambda_{PET} \mathcal{L}_{PET} + \lambda_{VQ} \mathcal{L}_{VQ} + \lambda_{SF} \mathcal{L}_{SF} + \lambda_{edge} \mathcal{L}_{edge}. \quad (15)$$

Here, $\mathcal{L}_{PET}$ is the PET reconstruction loss, $\mathcal{L}_{VQ}$ is the vector-quantization commitment and codebook loss [10], $\mathcal{L}_{SF}$ is the source-factorization loss over the shared, source-specific, and cross-source complementary latent partitions, and $\mathcal{L}_{edge}$ regularizes the structural conditioning pathway. The coefficients λ_{PET} , λ_{VQ} , λ_{SF} , and λ_{edge} balance the corresponding objective terms.

3 Experiments

3.1 Data and Preprocessing

We evaluated SFL-Net using ADNI-3 [25] and OASIS-3 [26] datasets. After preprocessing and quality control, 605 subjects were used for training and 83 held-out subjects were used for validation. Training and validation subjects were randomly drawn from both datasets. Subject metadata, including site, scanner, acquisition protocol, demographic variables, and cohort labels were not used for model training or evaluation.

For each subject, FLAIR MRI and tau-PET were registered to the corresponding T1-weighted MRI space, with a maximum acquisition time difference of 1.5 years between imaging modalities. Volumes were skull stripped using FreeSurfer SynthStrip [27–29], denoised, cropped, resized, and intensity normalized before training. PET-derived SUVR maps were computed by normalizing tau-PET uptake to cerebellar cortex uptake. Braak relevant regional masks were obtained using FreeSurfer SynthSeg [30] and grouped into Braak I/II (entorhinal cortex, parahippocampal gyrus, and hippocampus), Braak III/IV (limbic and temporal cortices), and Braak V/VI (parietal, frontal, and occipital cortices).

Subjects were assigned to five ordered tau stages using a sequential regional SUVR rule [31]:

$$\text{Stage} = \begin{cases} 4, & \text{SUVR}_{\text{braak V/VI}} > 1.873, \\ 3, & \text{SUVR}_{\text{braak III/IV}} > 1.523, \\ 2, & \text{SUVR}_{\text{braak III/IV}} > 1.307, \\ 1, & \text{SUVR}_{\text{braak I/II}} > 1.129, \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

The validation set contained 16 Stage 0, 45 Stage 1, 7 Stage 2, 10 Stage 3, and 5 Stage 4 subjects. Training used on-the-fly spatial augmentations, including random affine transformations, random axis flipping, and random resized cropping.

3.2 Baselines and Metrics

We compared SFL-Net against VAE, VQ-VAE, UNet, and Spatially-adaptive denormalization (SPADE) UNet®

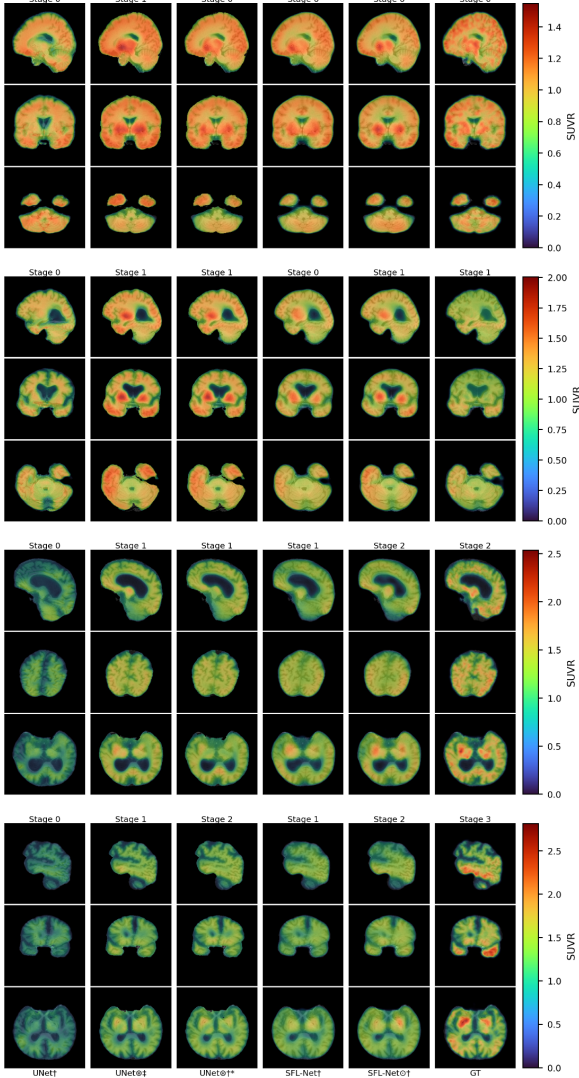


Figure 2: Representative SUVR reconstructions comparing ground truth, SFL-Net, and benchmark models.

baselines trained under matched input-output conditions [8, 10, 12, 32]. All models used the same MKConv backbone where applicable and were capacity matched to approximately 5 million learnable parameters. Training used an effective batch size of 8 with gradient accumulation, with 64 training cases randomly sampled per epoch. All models were trained for 1000 epochs using AdamW with cosine learning rate decay from 10^{-4} to 10^{-5} . Vector-quantized models used 20 warmup epochs with quantization disabled, followed by k -means codebook initialization and exponential moving average codebook updates.

Additional contrast-weighted and gradient-aware variants were included to evaluate whether conventional loss shaping improved high-uptake recovery independent of the source-factorized latent design [33–35]. We evaluated raw PET and SUVR MAE, MSE, PSNR, SSIM, MS-SSIM [36], and relative regional SUVR error, defined as the abso-

lute regional SUVR disagreement as a percentage of the true regional SUVR. We also evaluated high-uptake Dice score, sensitivity, specificity, regional Bland-Altman agreement [37], Braak-stage exact accuracy, within-one-stage accuracy, mean absolute stage error (MASE), and quadratic weighted kappa (QWK).

Continuous paired metrics were tested using two-sided Wilcoxon signed-rank tests, paired binary staging metrics using exact McNemar tests, and latent Shapley contributions using one-sided Wilcoxon signed-rank tests. Non-inferiority analyses were treated as sensitivity analyses and were assessed using paired bootstrap 90% confidence intervals, equivalent to one-sided 5% non-inferiority testing.

4 Results

4.1 Integrated Quantitative Performance

Table 1 summarizes reconstruction fidelity, high-uptake agreement, regional SUVR bias, and braak staging performance. SFL-Net[†] achieved the best raw PET MAE, MSE, and PSNR, the lowest absolute SUVR Bland-Altman bias, the lowest MASE, and the highest QWK. However, the added SSIM, MS-SSIM, and regional SUVR relative-error results show that performance was metric dependent. UNet[®] variants achieved the strongest structural-similarity metrics, the lowest regional SUVR relative error, and the best high-uptake Dice, sensitivity, and exact stage accuracy. Thus, SFL-Net[†] was strongest for raw PET intensity fidelity, regional SUVR bias, and ordinal stage displacement, whereas SPADE-based UNet baselines remained highly competitive for SUVR-domain structural similarity, relative SUVR error, and threshold-dependent metrics.

Paired subject-level tests supported this interpretation. Adding the source-factorization term improved SFL-Net[†] over SFL-Net[†] on all raw PET reconstruction metrics, including SSIM and MS-SSIM (all $p \leq 0.0151$; raw MS-SSIM $p = 1.18 \times 10^{-5}$). SUVR-domain gains were more selective, with significant improvements in SSIM ($p = 1.74 \times 10^{-4}$) and MS-SSIM ($p = 8.34 \times 10^{-9}$), but not in MAE, MSE, PSNR, or regional SUVR relative error ($p = 0.1675$). Regional relative error was tested by averaging absolute percent SUVR error across ROIs within each subject, followed by paired Wilcoxon signed-rank testing across subjects.

Adding SPADE inside SFL-Net did not improve performance. SFL-Net[†] outperformed SFL-Net[†] on raw and SUVR reconstruction metrics, including raw MS-SSIM ($p = 1.14 \times 10^{-13}$) and SUVR MS-SSIM ($p = 1.67 \times 10^{-13}$), while regional SUVR relative error was not significantly different ($p = 0.5984$). Compared with the strongest conventional staging baseline, UNet[®]*, SFL-Net[†] improved raw PET MSE and PSNR ($p = 0.0110$ and $p = 0.0171$), but not raw SSIM or MS-SSIM ($p = 0.1131$ and $p = 0.1283$). SUVR SSIM and MS-SSIM favored UNet[®] ($p = 0.0424$ and $p = 0.0086$), and regional SUVR relative error was not significantly dif-

Table 1: Validation results for evaluated models. Raw PET and SUVR normalized PET are assessed using reconstruction metrics, while high-uptake agreement is summarized by subject-level Dice, sensitivity, and specificity. Bland-Altman (BA) bias is computed over pooled subject ROI SUVR pairs from the five-level sequential staging regions, with %BA denoting percent SUVR bias. SUVR RelErr denotes relative regional SUVR error. Staging performance is evaluated using the five-level sequential staging rule over all validation subjects and over the early-stage subset restricted to ground-truth Braak stages 0–2. Model superscripts are defined as †: MSE reconstruction loss, ‡: Voxelwise Contrast Weighted Charbonnier reconstruction loss, *: Voxelwise Contrast Weighted Gradient Edge reconstruction loss, ⊗: SPADE, ⊙: source-factorization loss, γ: ResNet 18 supervision.

Model	Raw PET					SUVR PET					High uptake			BA bias		Braak stage								
																All stages				Early stages 0-2				
	MAE↓	MSE↓	PSNR↑	SSIM↑	MS-SSIM↑	MAE↓	MSE↓	PSNR↑	SSIM↑	MS-SSIM↑	RelErr↓	Dice↑	Sens.↑	Spec.↑	[SUVR → 0 % → 0]	Exact↑	Within-1↑	MASE↓	QWK↑	Exact↑	Within-1↑	MASE↓	QWK↑	
VAE†	0.1243	0.0212	16.65	0.838	0.747	0.3684	0.0846	18.30	0.841	0.747	46.3	0.124	0.091	0.940	-0.288	-18.4	0.193	0.735	1.313	0.000	0.235	0.897	0.868	0.000
VQ-VAE†	0.1215	0.0210	16.91	0.841	0.747	0.3571	0.0784	18.41	0.844	0.751	45.3	0.159	0.120	0.925	-0.283	-18.1	0.193	0.747	1.301	0.006	0.235	0.912	0.853	0.031
UNet†	0.1201	0.0209	17.07	0.854	0.812	0.3528	0.0819	19.08	0.855	0.812	31.2	0.161	0.118	0.947	-0.286	-18.4	0.193	0.735	1.313	0.000	0.235	0.897	0.868	0.000
UNet†⊗	0.1072	0.0178	18.31	0.861	0.850	0.2914	0.0543	20.51	0.866	0.857	12.8	0.507	0.513	0.827	-0.132	-5.8	0.494	0.819	0.819	-0.067	0.603	0.985	0.412	0.108
UNet†*	0.4219	0.2472	7.52	0.688	0.002	0.9712	0.7751	10.05	0.729	0.020	73.0	0.037	0.026	0.809	-0.886	-68.2	0.193	0.735	1.313	0.000	0.235	0.897	0.868	0.000
UNet†‡	0.1066	0.0175	18.36	0.863	0.854	0.2924	0.0555	20.44	0.868	0.861	13.4	0.570	0.636	0.771	-0.106	-3.5	0.530	0.831	0.735	-0.051	0.647	1.000	0.353	0.135
UNet†⊙	0.1072	0.0178	18.29	0.863	0.850	0.2938	0.0560	20.43	0.868	0.857	13.7	0.561	0.632	0.758	-0.102	-3.1	0.554	0.831	0.723	-0.002	0.676	0.985	0.338	0.165
UNet†*	0.1063	0.0175	18.37	0.862	0.850	0.2921	0.0555	20.48	0.867	0.857	13.2	0.542	0.574	0.796	-0.136	-5.9	0.518	0.831	0.771	-0.050	0.632	1.000	0.368	0.134
SFL-Net†	0.1085	0.0178	18.22	0.861	0.842	0.2958	0.0567	20.57	0.865	0.848	15.7	0.503	0.519	0.806	-0.146	-6.8	0.482	0.807	0.843	0.007	0.588	0.985	0.426	0.189
SFL-Net†γ	0.1135	0.0188	17.65	0.852	0.811	0.3252	0.0715	19.70	0.855	0.810	44.6	0.171	0.118	0.962	-0.236	-14.2	0.217	0.735	1.289	-0.019	0.265	0.897	0.838	-0.013
SFL-Net†⊗γ	0.1093	0.0181	18.13	0.858	0.837	0.3033	0.0604	20.37	0.861	0.840	26.6	0.378	0.320	0.904	-0.196	-11.0	0.277	0.771	1.133	0.005	0.338	0.941	0.721	-0.003
SFL-Net†‡	0.1043	0.0168	18.53	0.862	0.848	0.2918	0.0538	20.54	0.866	0.855	13.1	0.539	0.605	0.754	-0.101	-3.3	0.518	0.831	0.711	0.140	0.632	0.985	0.382	0.216
SFL-Net†‡	0.1088	0.0182	18.06	0.855	0.835	0.3044	0.0625	20.14	0.859	0.839	25.6	0.391	0.353	0.882	-0.192	-10.6	0.325	0.783	1.060	0.011	0.397	0.956	0.647	0.032
SFL-Net†⊙	0.1093	0.0181	18.12	0.856	0.831	0.3016	0.0619	20.29	0.860	0.837	16.6	0.430	0.399	0.868	-0.161	-8.0	0.422	0.807	0.928	-0.042	0.515	0.985	0.500	0.098
SFL-Net†⊙‡	0.1084	0.0180	18.12	0.857	0.839	0.3074	0.0641	20.14	0.860	0.843	23.1	0.430	0.406	0.862	-0.180	-9.4	0.361	0.795	1.024	-0.040	0.441	0.971	0.588	0.069

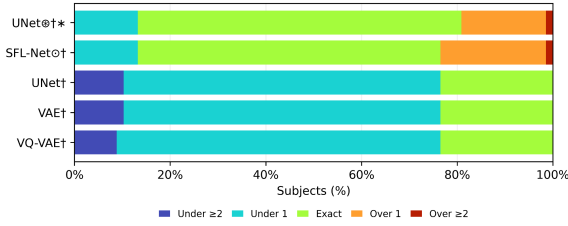


Figure 3: Signed early braak stage (0-2) error composition across synthesis models. Bars show the fraction of subjects with severe underestimation, one stage underestimation, exact agreement, one stage overestimation, or severe overestimation.

ferent ($p = 0.4456$). High-uptake Dice also favored UNet†* ($p = 0.0095$), whereas sensitivity, specificity, and downstream staging metrics were not significantly different.

4.2 Bias and braak Stage Error

Regional agreement was assessed with pooled Bland-Altman analysis across subject ROI SUVR pairs. SFL-Net†‡ had the smallest absolute SUVR bias (-0.101; 95% LoA [-0.805, 0.604]), while UNet†* had slightly smaller percent bias (-3.1%; 95% LoA [-39.4, 33.1]%).

For braak stage tracking, SFL-Net†‡ achieved the lowest mean absolute stage error (0.711) and highest QWK (0.140), whereas UNet†* achieved the highest exact stage accuracy (0.554). Overall, SFL-Net†‡ remains close or better in both all stage and early stage tracking metrics compared to the top baseline models Table. 1. The signed stage error composition in Fig. 3 shows that the best performing models concentrate more subjects in the exact

and one-stage error categories, whereas weaker baselines are dominated by larger underestimates, highlighting the design superiority of SFL-Net and dynamic loss based SPADE-UNet baselines.

Table 2: Paired non-inferiority sensitivity analysis for braak stage metrics. Differences are computed as SFL-Net†‡ minus UNet†*. CIs are paired bootstrap 90% confidence intervals. For exact and within-one-stage accuracy, values are in percentage points; for MASE, values are in braak-stage units. NI denotes non-inferiority established; Inc. denotes inconclusive non-inferiority.

	Exact acc. ↑	Within-1 acc. ↑	MASE ↓
Mean diff.	-3.6 pp	0.0 pp	-0.012 stages
LCI	-10.8 pp	0.0 pp	-0.096 stages
UCI	3.6 pp	0.0 pp	0.072 stages
Margin	-5 pp	-2.5 pp	+0.05 stages
Result	Inc.	NI	Inc.
Margin	-7.5 pp	-5 pp	+0.10 stages
Result	Inc.	NI	NI
Margin	-10 pp	-10 pp	+0.15 stages
Result	Inc.	NI	NI
Margin	-12.5 pp	-	+0.25 stages
Result	NI	-	NI

Table 2 summarizes paired non-inferiority sensitivity analyses against UNet†*. SFL-Net†‡ had slightly lower MASE than the baseline and satisfied non-inferiority for margins of +0.10 stages or larger. Within-one-stage accuracy was identical between models and remained non-inferior across all tested margins. Exact stage accuracy was slightly lower for SFL-Net†‡; non-inferiority was inconclusive for margins up to -10 percentage points and was established only under the more permissive -12.5 percentage-point margin. Thus, SFL-Net showed comparable ordinal-stage displacement and within-one-stage tolerance in this sensitivity analysis, while exact threshold-level staging remained the most margin-sensitive metric.

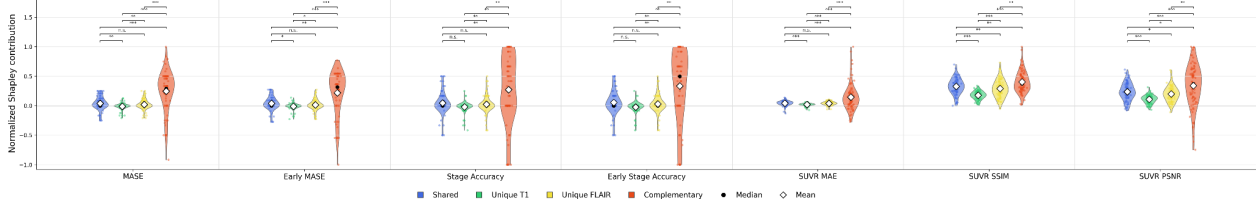


Figure 4: Latent-component Shapley attribution for SFL-Net using mean-replacement ablation. Significance bars denote pairwise comparisons between latent-component Shapley contributions within each metric using paired Wilcoxon signed-rank tests with Holm correction. *: $p_{\text{adj}} < 0.05$, **: $p_{\text{adj}} < 0.01$, ***: $p_{\text{adj}} < 0.001$; n.s. : non – significant. Early-stage metrics were computed for validation subjects with ground-truth Braak stage ≤ 2 . Error metrics were sign-adjusted such that positive contributions indicate lower error. Structure-only decoding served as the empty-coalition baseline for Shapley attribution.

4.3 Latent Attribution

We quantified the contribution of SFL-Net latent components using coalitional Shapley attribution over shared, T1-unique, FLAIR-unique, and cross-source complementary latent components [38–40]. This extends previous Shapley-based analysis of multi-contrast segmentation from input attribution to latent-component attribution in image synthesis [40]. Structure-only decoding served as the empty-coalition baseline, and missing latent components were ablated using mean replacement.

The cross-source complementary component produced the largest Shapley contribution across staging and SUVR-based reconstruction metrics, indicating that information jointly derived from both MRI contrasts was most important for downstream decoding performance (Fig. 4). Shared and FLAIR-unique components showed intermediate positive contributions, with their relative ordering depending on the metric. In contrast, the T1-unique component contributed weakly and was near-zero or negative for several staging-based metrics. Pairwise paired Wilcoxon signed-rank tests with Holm correction showed that the complementary component was significantly different from the other latent components across all evaluated metrics, while shared and FLAIR-unique contributions were not consistently separable for MASE, early MASE, stage accuracy, or SUVR MAE. These findings support the intended architectural role of the cross-source complementary stream and suggest that shared and FLAIR-unique information provide secondary but useful contributions, while T1-unique information is less consistently beneficial.

5 Discussion

In this study, we proposed and evaluated SFL-Net, a source-factorized and quantized multi-input representation for MRI to Tau-PET synthesis. The central contribution is an architecture that organizes the latent bottleneck into shared, source specific, and cross-source complementary pathways that can be interrogated through component ablation and Shapley attribution. Within the held-out validation set, the

strongest SFL-Net variant achieved competitive clinical and reconstruction performance among the evaluated models. These results were obtained without full encoder to decoder feature bypass, allowing latent components to be evaluated through coalition-based attribution rather than treating multi-contrast structural MRI to Tau-PET synthesis as a single opaque regression problem.

5.1 Multi-contrast structural MRI to Tau-PET synthesis is feasible

We show that T1 and FLAIR based multi-contrast structural MRI to Tau-PET synthesis is possible, although it remains a weakly constrained regression problem. T1 and FLAIR MRI do not directly measure tau binding, yet the evaluated models recovered clinically meaningful Tau-PET metrics, including regional SUVR agreement, high-uptake Tau overlap, Bland-Altman bias, Braak-stage agreement, and early staging performance. This supports the view that multi-contrast structural MRI contains disease related information that can be used to approximate Tau-PET signals.

We also extend existing structural MRI to Tau-PET synthesis work by evaluating stronger capacity-matched baselines in the same regression setting. Rather than comparing SFL-Net only against conventional autoencoder or simple UNet models, we included SPADE-based UNet variants and contrast/gradient-edge aware reconstruction objectives. The results show that these additions substantially improve conventional UNet performance. Compared with UNet[†], SPADE-based UNet models achieved much better SUVR reconstruction, high-uptake agreement, regional relative SUVR error, and Braak-stage metrics. The best SPADE-UNet variants also improved clinically oriented metrics, including high-uptake Dice and sensitivity, exact stage accuracy, within-one-stage agreement, and early-stage accuracy. Thus, the baseline analysis shows that structural conditioning and adaptive reconstruction losses can improve both quantitative reconstruction and clinically relevant downstream metrics.

Against these stronger baselines, our proposed model, SFL-Net^{⊕†}, achieved similar top-tier performance across recon-

struction and clinical benchmarks. It achieved the best raw PET MAE, MSE, and PSNR, the lowest absolute regional SUVR Bland-Altman bias, the lowest MASE, and the highest QWK among the evaluated models. While not the best, our method remained competitive among several threshold and similarity dependent metrics, including SUVR SSIM/MS-SSIM, regional SUVR relative error, high-uptake Dice, and exact stage accuracy. The non-inferiority sensitivity analysis further showed that SFL-Net[†] preserved within-one-stage tolerance and comparable ordinal stage displacement relative to the strongest baseline, while exact stage accuracy remained margin sensitive. Thus, SFL-Net shows that our proposed method maintains competitive clinical performance while adding an auditable latent bottleneck that conventional UNet based models lack.

The ablation behavior also explains why SPADE and adaptive reconstruction losses did not improve SFL-Net in the same way they improved the UNet baselines. For the UNet baselines, SPADE adds valuable ROI-level spatial conditioning information using semantic segmentation maps of the brain, leading to improved signal recovery by ROI mask. In SFL-Net, however, anatomical preservation is already encouraged by the source-factorized bottleneck and structure-conditioned latent decoder. Adding SPADE to this architecture degraded performance, suggesting that the existing factorized latent design may already preserve anatomical information efficiently and that additional spatial modulation can interfere with the intended bottleneck organization. Similarly, adaptive contrast aware objectives worsened SFL-Net performance in this implementation. Based on training behavior, this likely reflects reduced stability of the vector-quantized codebooks under these objectives rather than a general failure of adaptive losses. These results suggest that objectives and conditioning mechanisms that benefit continuous UNet-based regression pipelines may not transfer directly to a quantized source-factorized architecture without additional stabilization.

Overall, SFL-Net is best suited for synthesis settings where UNet derived skip connections may be a pitfall for downstream tasks. Skip connections can preserve fine spatial detail, but they also create high-bandwidth routes around the bottleneck, weakening the ability to attribute the output to specific latent mechanisms. SFL-Net achieves comparable clinical and reconstruction performance without full encoder-decoder feature bypass, making it a more appropriate design when interpretability, bottleneck accountability, and source-level auditing are central requirements.

5.2 Latent source factorization provides component-level audibility

The latent Shapley analysis shows that SFL-Net effectively separates source-factorized latent components into functionally specialized partitions. This separation is encouraged by the architecture itself, which assigns shared, T1-unique, FLAIR-unique, and cross-source complementary

pathways, and is reinforced by the source-factorization loss. Across most clinical and reconstruction metrics, component-level importance followed the same broad ordering: complementary > shared > FLAIR-unique > T1-unique. This pattern indicates that the most useful information for decoding Tau-PET signal is not contained in either structural MRI contrast alone, but in the joint configuration of T1-weighted and FLAIR MRI.

The dominant complementary contribution is important because it supports the intended role of the cross-source pathway. For MRI to Tau-PET synthesis, the model must infer a molecular imaging target from structural and disease associated anatomical cues. A strong complementary component suggests that the prediction benefits from interaction-dependent information across T1-weighted and FLAIR MRI, such as the spatial relationship between anatomy, atrophy patterns, tissue abnormalities, and regional disease context. This suggests that the multi-contrast input configuration is being used in a nontrivial way.

The relative ordering of FLAIR-unique and T1-unique information is also informative. The stronger FLAIR-unique contribution suggests that after shared anatomy has been captured, residual FLAIR-specific signal may provide more disease-relevant context than residual T1-specific signal. By contrast, much of the useful T1-weighted information may already be absorbed into the shared anatomical and structural-conditioning pathways, leaving comparatively little residual T1-unique information relevant for decoding Tau signals.

This source-factorized interpretation is the central practical advantage of SFL-Net. It allows the synthesis process to be audited by asking whether the generated Tau-PET signal is driven primarily by shared anatomy, by one MRI contrast, or by cross-source interactions. This is more informative than treating the synthesis model as a single monolithic regressor. The Shapley results therefore help rationalize the model output and provide evidence that the latent partitions carry specialized information.

5.3 Clinical metrics and early-stage evaluation are necessary

The results also show that reconstruction metrics alone do not fully characterize clinical utility. SFL-Net[†] was strongest for raw PET intensity fidelity and regional signed bias, whereas SPADE-UNet variants were stronger for SUVR structural similarity, high-uptake Dice, and exact stage accuracy. These discrepancies are expected because voxelwise image similarity, regional SUVR calibration, high-uptake localization, and threshold-based Braak staging measure different properties of the prediction. A model can have favorable global reconstruction error while still missing sparse high-uptake regions, and a model can have low signed Bland-Altman bias while retaining substantial absolute regional error because positive and negative errors cancel across ROIs and subjects. Therefore, evaluation of synthetic Tau-PET should include clinically oriented

metrics such as regional SUVR error, high-uptake overlap, signed and absolute bias, exact and within-one-stage agreement, MASE, QWK, and early-stage performance.

Early-stage evaluation is particularly important for this application as the most relevant use case is not simply recognizing advanced disease after substantial neurodegeneration is already visible on structural MRI, but distinguishing low burden or early tau-positive patterns and supporting early disease tracking. Later stage disease may be accompanied by more obvious atrophy or structural abnormality, whereas early stage changes are subtle and more clinically important for distinguishing cognitively normal and mild cognitive impairment, or evaluating disease trajectory. For this reason, early stage performance is a more appropriate and clinically relevant test for MRI to Tau-PET synthesis.

Additionally, the validation set contained many more Stage 0 and Stage 1 subjects than Stage 3 and Stage 4 subjects, making higher-stage predictions underperform. In addition, Braak-stage assignment depends on regional SUVR thresholds, so small reconstruction errors, segmentation uncertainty, or registration misalignment can alter the predicted stage. These issues are amplified when the decisive uptake pattern is spatially sparse or heterogeneous. Together, the class imbalance and threshold sensitivity help explain the observed mean-pulling behavior, where most models tend to underestimate higher stages and pull predictions toward the dominant lower-stage training distribution. Thus, whole-stage results should be interpreted together with early-stage metrics, signed-error composition, Bland-Altman analysis, and regional SUVR errors.

6 Conclusion

SFL-Net offers a practical alternative to UNet-based MRI to Tau-PET synthesis in settings where interpretability, bottleneck transparency, and source-level auditing are prioritized over reconstruction accuracy alone. The model delivers competitive clinical and reconstruction performance without relying on full encoder-decoder skip connections, and its source-factorized latent architecture supports component-wise attribution of shared, T1-specific, FLAIR-specific, and cross-source complementary contributions. While SFL-Net is not a direct substitute for stronger UNet baselines, it attains similarly competitive performance with an auditable bottleneck that clarifies how multi-contrast structural MRI inputs shape the generated Tau-PET signal.

Acknowledgment

This work was supported in part by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research, DOE Computational Science Graduate Fellowship under Award No. DE-SC0024386 to Juampablo E. Heras Rivera.

References

- [1] C. R. Jack, D. A. Bennett, K. Blennow *et al.*, “Nia-aa research framework: Toward a biological definition of alzheimer’s disease,” *Alzheimer’s & Dementia*, vol. 14, no. 4, pp. 535–562, 2018.
- [2] S. C. Burnham, M. D. Devous *et al.*, “A review of the flortaucipir literature for positron emission tomography imaging of tau neurofibrillary tangles,” *Brain Communications*, vol. 6, no. 1, p. fcad305, 2024.
- [3] S.-D. Chen *et al.*, “Staging tau pathology with tau pet in alzheimer’s disease: A longitudinal study,” *Translational Psychiatry*, 2021.
- [4] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, “The clinical use of structural mri in alzheimer disease,” *Nature Reviews Neurology*, vol. 6, no. 2, pp. 67–77, 2010.
- [5] S. Dayarathna *et al.*, “Deep learning based synthesis of mri, ct and pet: Review and analysis,” *Medical Image Analysis*, vol. 92, p. 103046, 2024.
- [6] J. Lee, B. J. Burkett, H.-K. Min *et al.*, “Synthesizing images of tau pathology from cross-modal neuroimaging using deep learning,” *Brain*, vol. 147, no. 3, pp. 980–995, 2024.
- [7] J. Moon, S. Kim, H. Chung, I. Jang, and A. D. N. Initiative, “Cyclic 2.5d perceptual loss for cross-modal 3d medical image synthesis: T1w mri to tau pet,” *Human Brain Mapping*, vol. 47, no. 5, p. e70508, 2026.
- [8] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *MICCAI*, 2015.
- [9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *CVPR*, 2017.
- [10] A. van den Oord, O. Vinyals, and K. Kavukcuoglu, “Neural discrete representation learning,” in *NeurIPS*, 2017.
- [11] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. Courville, “FiLM: Visual reasoning with a general conditioning layer,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
- [12] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, “Semantic image synthesis with spatially-adaptive normalization,” in *CVPR*, 2019.
- [13] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, “Edgeconnect: Structure guided image inpainting using edge prediction,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 3265–3274.
- [14] Y. Luo, D. Nie, B. Zhan, Z. Li, X. Wu, J. Zhou, Y. Wang, and D. Shen, “Edge-preserving MRI image synthesis via adversarial network with iterative multi-scale fusion,” *Neurocomputing*, vol. 452, pp. 63–77, 2021.
- [15] P. L. Williams and R. D. Beer, “Nonnegative decomposition of multivariate information,” *arXiv preprint arXiv:1004.2515*, 2010. [Online]. Available: <https://arxiv.org/abs/1004.2515>
- [16] V. Griffith and T. Ho, “Quantifying redundant information in predicting a target random variable,” *Entropy*, vol. 17, no. 7, pp. 4644–4653, 2015. [Online]. Available: <https://www.mdpi.com/1099-4300/17/7/4644>
- [17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 2006.
- [18] E. Jang, S. Gu, and B. Poole, “Categorical reparameterization with gumbel-softmax,” in *ICLR*, 2017.
- [19] C. J. Maddison, A. Mnih, and Y. W. Teh, “The concrete distribution: A continuous relaxation of discrete random variables,” in *ICLR*, 2017.
- [20] A. Strehl and J. Ghosh, “Cluster ensembles – a knowledge reuse framework for combining multiple partitions,” *Journal of Machine Learning Research*, vol. 3, pp. 583–617, 2002.
- [21] T. O. Kvålseth, “On normalized mutual information: Measure derivations and properties,” *Entropy*, vol. 19, no. 11, p. 631, 2017.
- [22] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, “Detecting novel associations in large data sets,” *Science*, vol. 334, no. 6062, pp. 1518–1524, 2011.
- [23] I. Sobel and G. Feldman, “A 3x3 isotropic gradient operator for image processing,” Stanford Artificial Intelligence Project, Tech. Rep., 1968.
- [24] J. Abderezaei, A. Pionteck, A. Chopra, and M. Kurt, “3d inception-based transmorph: Pre- and post-operative multi-contrast mri registration in brain tumors,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, S. Bakas, A. Crimi, U. Baid, S. Malec, M. Pytlaiz, B. Baheti, M. Zenk, and R. Dorent, Eds. Cham: Springer Nature Switzerland, 2023, pp. 35–45.
- [25] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. J. Jack, W. Jagust, J. C. Morris, R. C. Petersen, J. Salazar, A. J. Saykin, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and A. D. N. Initiative, “The alzheimer’s disease neuroimaging initiative 3: Continued innovation for clinical trial improvement,” *Alzheimer’s & Dementia*, vol. 13, no. 5, pp. 561–571, May 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.jalz.2016.10.006>
- [26] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hasenstab, K. Moulder, A. G. Vlassenko, M. E. Raichle, C. Cruchaga, and D. Marcus, “Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for

- normal aging and alzheimer disease,” *bioRxiv*, 2019. [Online]. Available: <https://www.medrxiv.org/content/10.1101/2019.12.13.19014902v1>
- [27] A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, and M. Hoffmann, “SynthStrip: skull-stripping for any brain image,” *NeuroImage*, vol. 260, p. 119474, 2022.
 - [28] W. Kelley, N. Ngo, A. V. Dalca, B. Fischl, L. Zöllei, and M. Hoffmann, “Boosting skull-stripping performance for pediatric brain images,” in *IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2024, pp. 1–5.
 - [29] M. Hoffmann, “Domain-randomized deep learning for neuroimage analysis,” *IEEE Signal Processing Magazine*, vol. 42, no. 4, pp. 78–90, 2025.
 - [30] B. Billot, D. N. Greve, O. Puonti, A. Thielscher, K. Van Leemput, B. Fischl, A. V. Dalca, and J. E. Iglesias, “Synthseg: Segmentation of brain mri scans of any contrast and resolution without retraining,” *Medical Image Analysis*, vol. 86, p. 102789, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1361841523000506>
 - [31] S.-D. Chen, J.-Y. Lu, H.-Q. Li, Y.-X. Yang, J.-H. Jiang, M. Cui, C.-T. Zuo, L. Tan, Q. Dong, J.-T. Yu, and A. D. N. Initiative, “Staging tau pathology with tau pet in alzheimer’s disease: a longitudinal study,” *Translational Psychiatry*, vol. 11, no. 1, p. 483, 2021. [Online]. Available: <https://doi.org/10.1038/s41398-021-01602-5>
 - [32] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” in *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014. [Online]. Available: <https://openreview.net/forum?id=33X9fd2-9FyZd>
 - [33] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, “Two deterministic half-quadratic regularization algorithms for computed imaging,” *ICIP*, 1994.
 - [34] M. Mathieu, C. Couprie, and Y. LeCun, “Deep multi-scale video prediction beyond mean square error,” in *ICLR*, 2016.
 - [35] J. Abderezaei, A. Pionteck, A. Chopra, and M. Kurt, “3d inception-based transmorph: Pre- and post-operative multi-contrast mri registration in brain tumors,” in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2022. Lecture Notes in Computer Science*. Springer, Cham, February 5 2024, vol. 14092, pp. 35–45.
 - [36] A. Horé and D. Ziou, “Image quality metrics: Psnr vs. ssim,” in *2010 20th International Conference on Pattern Recognition*, 2010, pp. 2366–2369.
 - [37] J. M. Bland and D. G. Altman, “Statistical methods for assessing agreement between two methods of clinical measurement,” *The Lancet*, vol. 327, no. 8476, pp. 307–310, 1986.
 - [38] L. S. Shapley, “A value for n-person games,” in *Contributions to the Theory of Games II*, H. W. Kuhn and A. W. Tucker, Eds. Princeton University Press, 1953, pp. 307–317.
 - [39] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
 - [40] T. Ren, J. Heras Rivera, H. Oswal, Y. Pan, A. Chopra, J. Ruzevick, and M. Kurt, “Here comes the explanation: A shapley perspective on multi-contrast medical image segmentation,” in *Joint Proceedings of the xAI 2025 Late-breaking Work, Demos and Doctoral Consortium co-located with the 3rd World Conference on eXplainable Artificial Intelligence (xAI 2025)*, ser. CEUR Workshop Proceedings, vol. 4017. Istanbul, Turkey: CEUR-WS.org, 2025, pp. 185–192. [Online]. Available: <https://ceur-ws.org/Vol-4017/paper.24.pdf>