MICCAI2025
28th International Conference on Medical Image Computing and Computer Assisted Intervention
23-27 September 2025
Daejeon Convention Center
Daejeon REPUBLIC OF KOREA

Spherical Diffusion Process for Score-guided Cortical Correspondence via Spectral Attention

Seungeun Lee[1] Sergey Pyatkovskiy[2] Jaejun Yoo[2] Ilwoo Lyu[3]

[1] Klleon [2] Ulsan National Institue of Science and Technology [3] Pohang University of Science and Technology

Objective

To align cortical geometrc features (e.g. sulcal depth) across subjects by estimating warp field Φ so that at each spherical location (θ,φ) on the target, the target feature $F(\theta,\varphi)$ should match the source M sampled at the warped location $\Phi(\theta,\varphi)$.

Motivation

- **High anatomical variability across** individuals hampers robust one-shot alignment.
- **Non-Euclidean geometry** of spherical maps calls for a manifold-aware generative prior.
- **Efficiency & robustness**: Spatial attention over ~40k vertices is prohibitively expensive; spectral operations scale better.

Contributions

- **Score on $\mathbb{S}^2$**: Formulate diffusion process with the Laplace-Beltrami operator in spherical harmonics (SH), enabling closed-form kernels on $\mathbb{S}^2$.
- **Conditional guidance:** Learn a conditional score that models transition from source to target, to guide warp field estimation.
- **Spectral cross-attention:** Condition deformation in SH with score cues for stable, low-distortion alignment.

Method

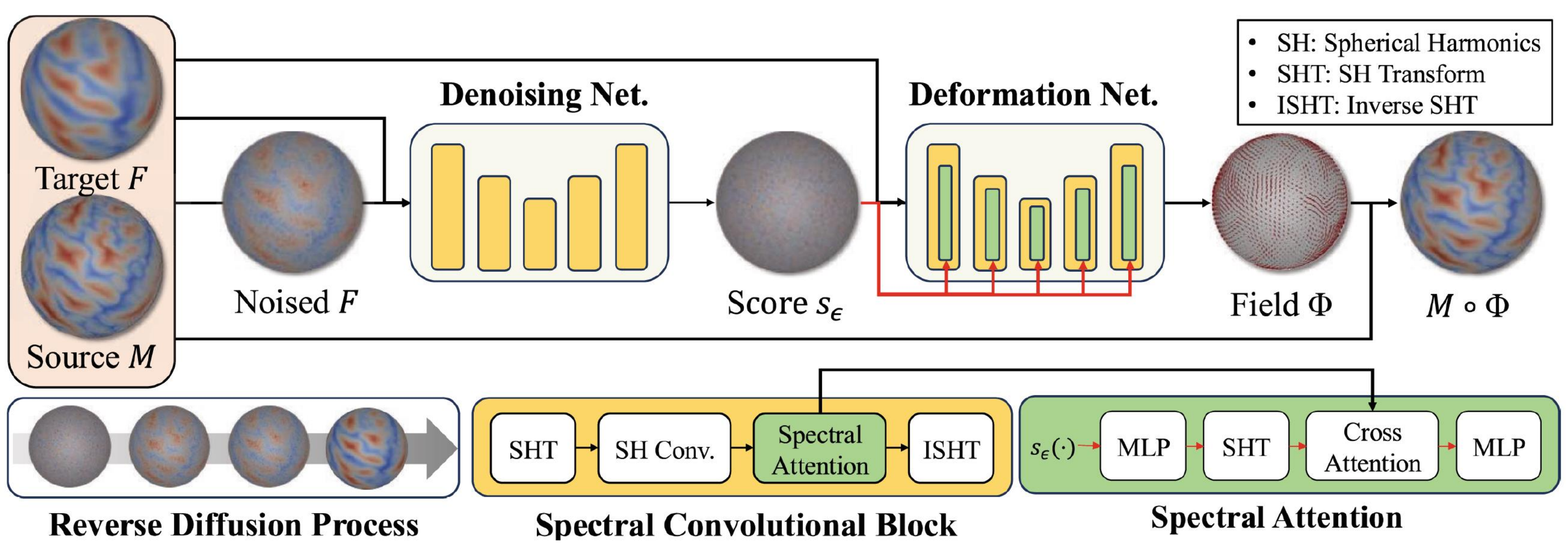

Fig. 1. Overall pipeline of our proposed method.

Denoising network: Spherical harmonics-based network estimates the conditional score in SH space at each diffusion time.

Deformation network takes M, F, and the final-step score embedding; integrated spectral cross-attention at warp module layer to produce Φ.

Joint training: Optimize score matching + similarity + regularization.

❖ Spherical Diffusion Process

Core idea. On $\mathbb{S}^2$, cortical maps are corrupted by spherical heat diffusion; we learn a conditional score in the spherical-harmonic domain. At inference, reverse integration denoises toward the target, and the score embedding conditions a deformation net via spectral attention for low-distortion correspondence.

❖ Spectral Attention

Benefit: Complexity depends on SH bandwidth L (e.g. $L = 40$), avoiding the quadratic cost over 40k spatial vertices.

Compute cross-attention in SH domain between the score embedding and intermediate deformation features; Queries from deformation features, key/values from score embedding.

Experimental Setup

- **Dataset:** HCP (1,113 scans, train); Mindboggle (101 scans, test).
- **Comparative Methods:** Optimization-based (FS, SD, MSM, HSD), and learning-based method (Lee *et al.,* SHAPEMI'24).
- **Statistical computation:** paired t-test after FDR ($q < 0.05$).

Quantitative Results

Method	Accuracy			Areal Distortion			Edge Distortion		
	MSE↓	NCC↑	Dice↑	Mean↓	Median↓	P_{95} ↓	Mean↓	Median↓	P_{95} ↓
FS	0.313	0.898	0.874	0.322	0.267	0.792	0.136	0.109	0.349
SD	0.309	0.898	0.873	0.320	0.250	0.861	0.153	0.117	0.418
MSM	0.311	0.891	0.863	0.573	0.404	1.703	0.257	0.196	0.708
HSD	0.305	0.898	0.872	0.306	0.234	0.830	0.146	0.116	0.385
Lee *et al.*	0.307	0.899	0.871	0.289	0.225	0.775	0.144	0.112	0.384
Ours	**0.294**	**0.910**	**0.880**	**0.275**	**0.204**	**0.750**	**0.131**	**0.103**	**0.349**

Table 1. Quantitative comparison on HCP (blue: statistical improvement).

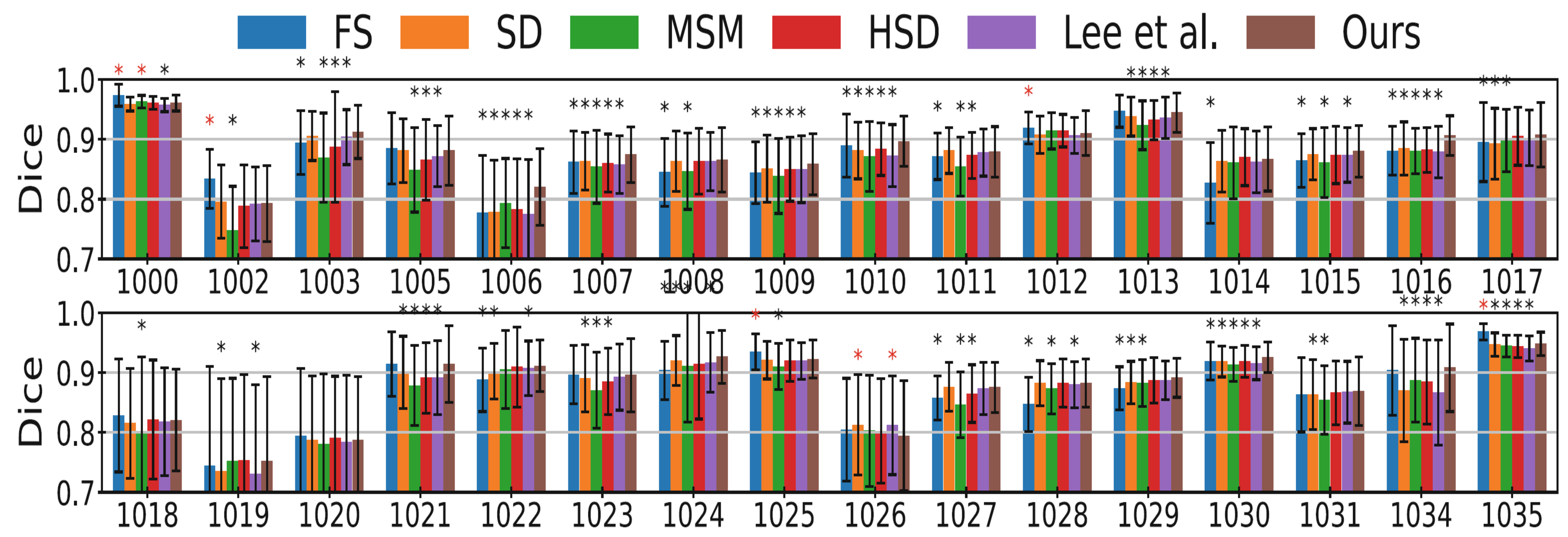

Fig. 2. 32 Cortical Region-wise dice score comparison on Mindboggle-101.

- Our method achieves the best accuracy with the lowest areal and edge distortions among all baselines under matched accuracy.
- Under matched registration accuracy (~0.3 MSE), our method shows significantly lower shape distortions than all baselines.
- Significant Dice improvements across most of the 32 cortical regions versus the baselines.

Qualitative Results

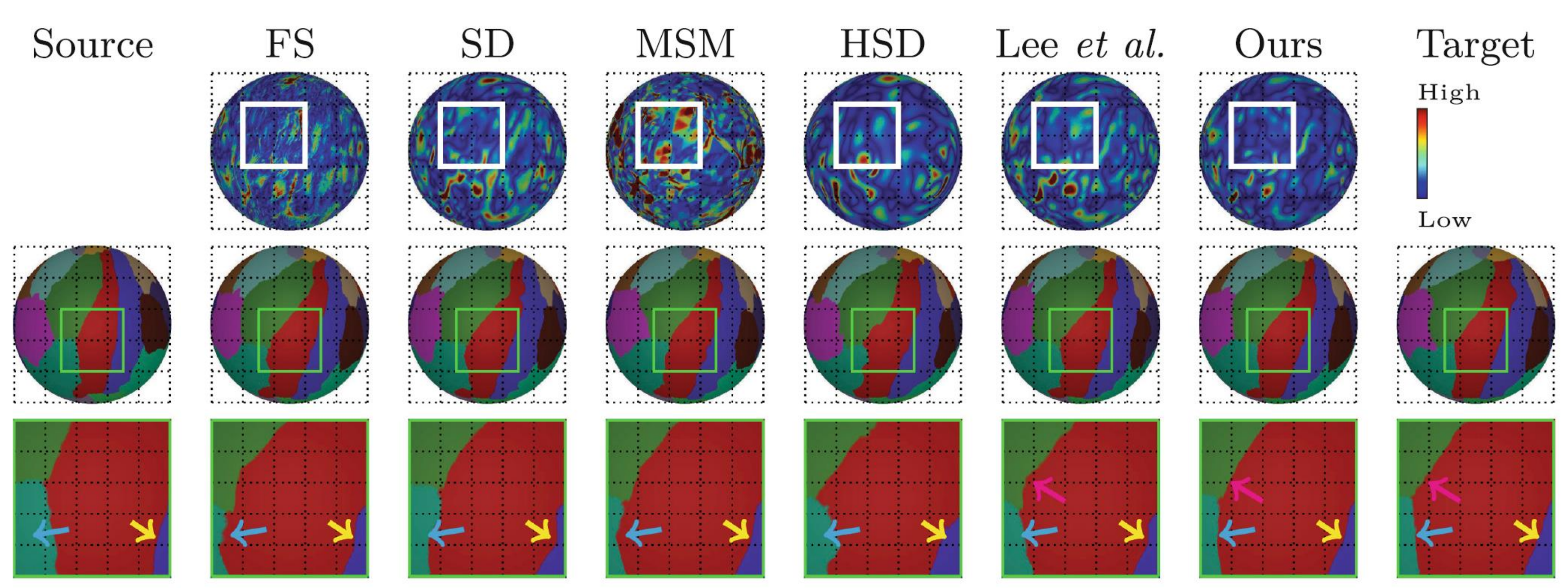

Fig. 3. Visual inspection of an example subject (1st row: areal distortion, 2nd row: parcellation maps, and 3rd row: its cropped view

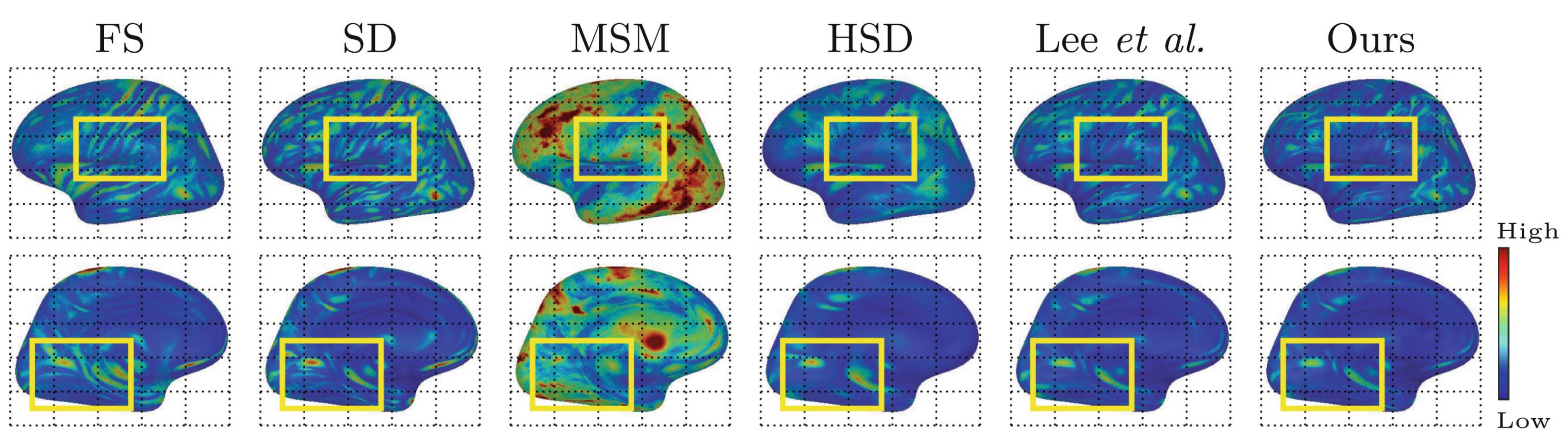

Fig. 4. The group average maps of areal distortion across participants.

- Visual example shows markedly reduced areal distortion (e.g., supramarginal region) and cleaner parcellation boundaries (e.g., postcentral area).
- Group-average areal distortion decreases broadly, especially in frontal view.