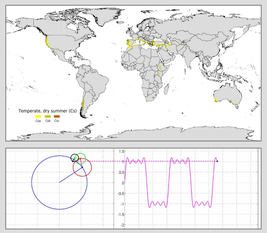
LocDiff: Identifying Locations on Earth by Diffusing in the Hilbert Space

Zhangyu Wang, Zeping Liu, Jielu Zhang, Zhongliang Zhou, Qian Cao, Nemin Wu, Lan Mu, Yang Song, Yiqun Xie, Ni Lao, Gengchen Mai

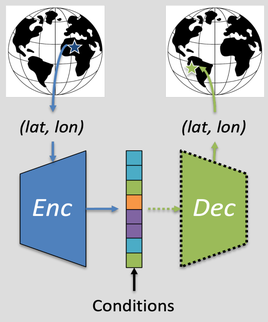
NEURAL INFORMATION PROCESSING SYSTEMS

Background

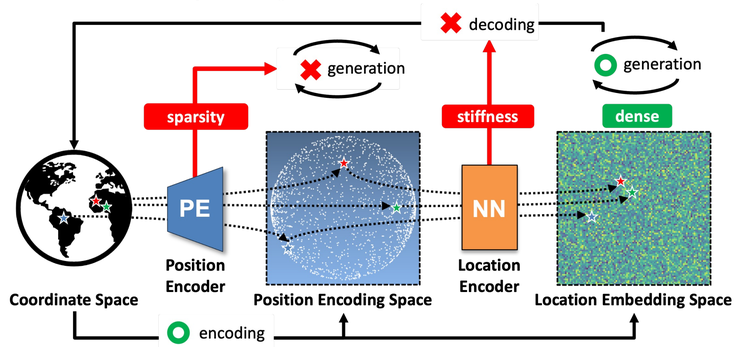
Multi-Scale Location Encoding



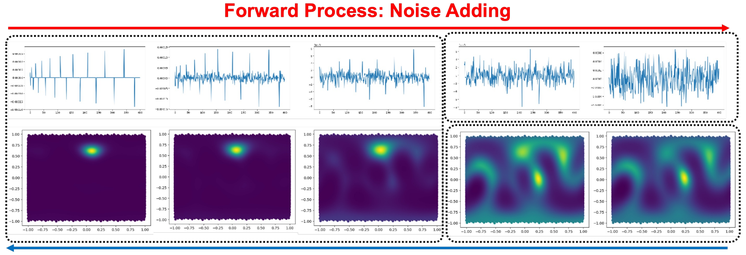
Conditional Location Decoding



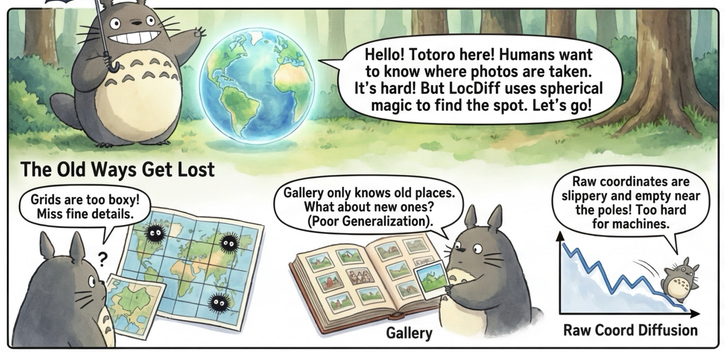
Motivation: Sparsity & Stiffness Problems

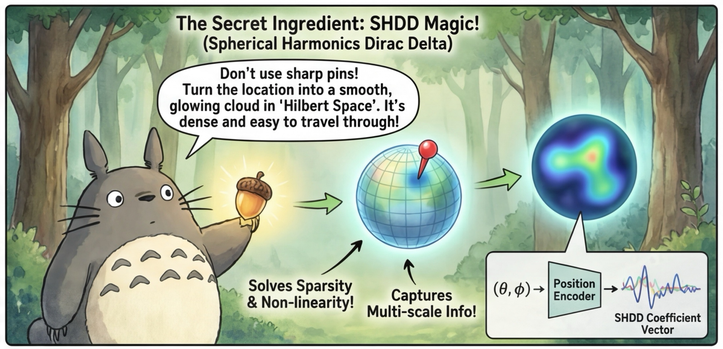


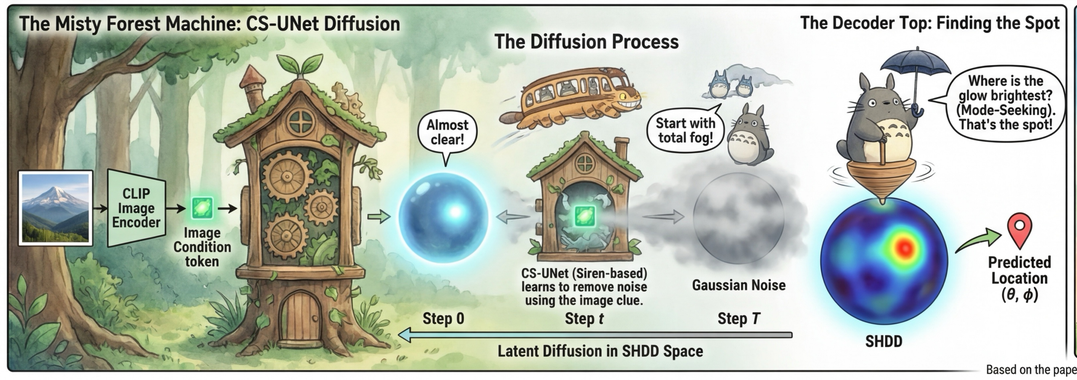
Goal: Latent (Location Encoding) Diffusion



Totoro's Guide to LocDiff: Finding Places with Hilbert Space Magic!





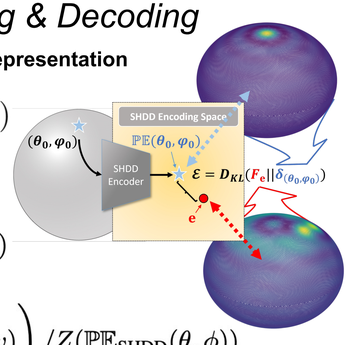


Based on the paper 'LocDiff: Identifying Locations on Earth by Diffusing in the Hilbert Space'.

Solution: SHDD Encoding, Learning & Decoding

Spherical Harmonics Dirac Delta-function (SHDD) Representation

$$\mathbb{PE}^{L}_{\text{SHDD}}(\theta_0, \phi_0) = \bigcup_{l=0}^{\infty} \bigcup_{m=-l}^{l} Y_{lm}(\theta_0, \phi_0)$$



SHDD Distributions

$$q_{\mathbf{e}}(u, v) := \exp\left(\sum_{l=0}^{L} \sum_{m=-l}^{l} e_{lm} Y_{lm}(u, v)\right) / Z(\mathbf{e})$$

$$p_{(\theta,\phi)}(u, v) := \exp\left(\sum_{l=0}^{L} \sum_{m=-l}^{l} Y_{lm}(\theta, \phi) Y_{lm}(u, v)\right) / Z(\mathbb{PE}_{\text{SHDD}}(\theta, \phi))$$

SHDD KL-divergence Loss Function

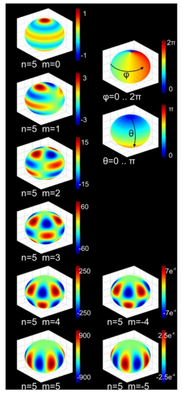
$$\mathcal{L}_{\text{SHDD-KL}}(\mathbf{e}, (\theta, \phi)) := \int_0^{\pi} \int_0^{2\pi} q_{\mathbf{e}}(u, v) \log \frac{q_{\mathbf{e}}(u, v)}{p_{(\theta,\phi)}(u, v)} \mathbf{d}u\mathbf{d}v$$

SHDD Mode-seeking Location Decoding

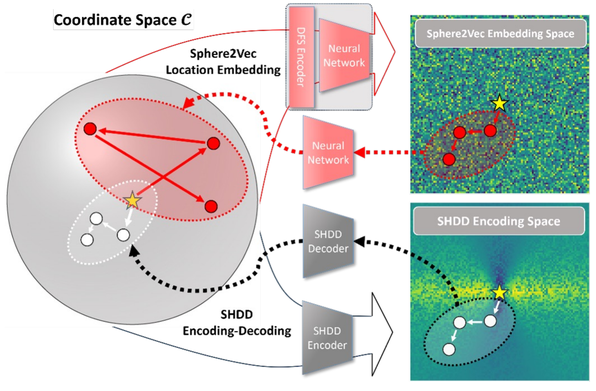
$$\mathbb{PD}_{\text{mode}}(\mathbf{e}; \rho) := \underset{(\theta,\phi)}{\arg\max} \left\{ \int_{\theta-\rho}^{\theta+\rho} \int_{\phi-\rho}^{\phi+\rho} \exp\left(\sum_{l=0}^{L} \sum_{m=-l}^{l} e_{lm} Y_{lm}(u, v)\right) \mathbf{d}u\mathbf{d}v \right\}$$

Advantages: Multi-Scale, Dense, Smooth & Stable

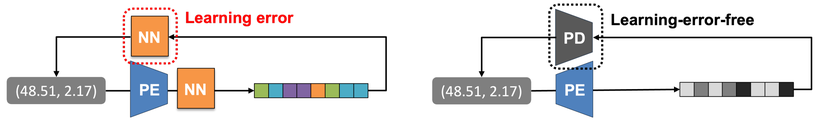
Spherical Harmonics



Dense & Smooth SHDD Encoding Space



Neural Encoding & Decoding v.s. SHDD Encoding & Decoding



Experimental Results

Geo-localization Accuracy Improvement

Model	Im2GPS3k					YFCC-26k					GWS15k				
	Street 1 km	City 25 km	Region 200 km	Coun. 750 km	Cont. 2.5k km	Street 1 km	City 25 km	Region 200 km	Coun. 750 km	Cont. 2.5k km	Street 1 km	City 25 km	Region 200 km	Coun. 750 km	Cont. 2.5k km
[L]kNN, σ=4 [46]	7.2	19.4	26.9	38.9	55.9	-	-	-	-	-	-	-	-	-	-
PlaNet [47]	8.5	24.8	34.3	48.4	64.6	4.4	11.0	16.9	28.5	47.7	-	-	-	-	-
CPlaNet [38]	10.2	26.5	34.6	48.6	64.6	-	-	-	-	-	-	-	-	-	-
ISNs [30]	10.5	28.0	36.6	49.7	66.0	5.3	12.3	19.0	31.9	50.7	0.05	0.6	4.2	15.5	38.5
Translocator [32]	11.8	31.1	46.7	58.9	80.1	7.2	17.8	28.0	41.3	60.6	0.5	1.1	8.0	25.5	48.3
GeoDecoder [6]	12.8	33.5	45.9	61.0	76.1	10.1	23.9	34.1	49.6	69	0.7	1.5	8.7	26.9	50.5
GeoCLIP [45]	14.1	34.5	50.7	69.7	83.8	11.6	22.2	36.7	57.5	76.0	0.6	3.1	16.9	45.7	74.1
PIGEON [10]	11.3	**36.7**	53.8	72.4	85.3	10.5	25.8	**42.7**	63.2	79.0	0.7	9.2	31.2	65.7	**85.1**
LocDiff (L=47)	10.9	34.0	53.3	72.5	85.2	9.6	22.8	37.5	58.6	76.8	**2.1**	**12.4**	**33.7**	**67.0**	85.0
LocDiff-H (L=23)	**15.3**	36.5	**56.4**	**75.2**	**87.4**	**13.2**	**26.0**	41.9	**64.5**	**80.3**	0.9	7.4	33.5	66.2	85.0

Spatial Generalizability Improvement

Model	Gallery/Anchor	Size	Street 1 km	City 25 km	Region 200 km	Country 750 km	Continent 2500 km
GeoCLIP	MP16	100 k	14.11	34.47	50.65	69.67	83.82
	Grid	1 M	0.03 (↓99.79%)	9.18 (↓73.37%)	33.47 (↓33.90%)	55.32 (↓20.63%)	75.34 (↓10.11%)
		500 k	0.03 (↓99.79%)	7.17 (↓79.21%)	29.40 (↓41.96%)	52.29 (↓24.94%)	73.11 (↓12.80%)
		100 k	0.00 (↓100.00%)	2.67 (↓92.25%)	22.39 (↓55.81%)	47.35 (↓32.05%)	68.77 (↓17.94%)
		21 k	0.00 (**↓100.00%**)	0.87 (**↓97.48%**)	19.55 (**↓61.41%**)	43.78 (**↓37.17%**)	64.33 (**↓23.26%**)
LocDiff (L=23)	MP16	100 k	0.57	11.1	44.42	68.35	82.50
	Grid	1 M	0.01 (↓98.25%)	4.37 (↓60.63%)	43.04 (↓3.10%)	68.30 (↓0.07%)	81.66 (↓1.02%)
		500 k	0.07 (↓87.72%)	4.47 (↓59.73%)	43.18 (↓2.79%)	68.36 (↑0.01%)	81.65 (↓1.03%)
		100 k	0.07 (↓87.72%)	4.04 (↓63.60%)	42.91 (↓3.40%)	68.34 (↓0.01%)	82.18 (↓0.39%)
		21 k	0.03 (↓94.74%)	4.90 (↓55.86%)	43.44 (↓2.21%)	68.29 (↓0.09%)	81.68 (↓0.99%)