How Should Geometry Guide Latent Reasoning? MeRa for Spatial Prediction

Anonymous authors
Paper under double-blind review

Abstract

Post-encoder refinement revisits an encoded history before prediction, but additional depth alone does not specify which relations guide that retrieval. For spatial tasks, this raises a specific question: how should iterative refinement use the geometry of the input? We study this question through **MeRa** (**Me**tric-space **Rea**soning), a modular post-encoder interface that combines distance-conditioned cross-attention, fixed encoder memory, and gated residual updates. The original experimental record reports higher NDCG@10 with MeRa in six backbone–dataset settings, with relative gains of 2.5–13.9% for GETNext and 5.8–27.7% for LSTM. On NYC GETNext, the reported gap between refinement without and with distance bias is 4.5%. We characterize finite-depth sensitivity and fixed-memory information access, clarifying what the architecture guarantees and what must be assessed empirically. Depth comparisons and a CLEVR-derived nearest-object task provide complementary observations. The results motivate geometry-conditioned refinement as a design choice; unresolved historical run alignment and evaluation provenance limit causal and cross-protocol performance claims. Code is available at https://anonymous.4open.science/r/MeRaa.

1 Introduction

Spatial prediction couples representation learning with the geometry of the prediction space. In next point-of-interest (POI) recommendation, a model predicts a user's next location from a check-in trajectory. A useful representation must summarize temporal preferences while retaining spatial relations among visited locations (Yang et al., 2022; Luo et al., 2021). A strong sequence encoder need not expose those relations in the form most useful to the final prediction head.

Recent recommendation systems improve sequence encoding through graph structure, geographic priors, and hyperbolic representations (Yang et al., 2022; Qin et al., 2025; Li et al., 2025). A complementary line of work studies latent reasoning: allocating additional computation to continuous representations before prediction (Tang et al., 2026b; Liu et al., 2025a; Tang et al., 2026a; Geiping et al., 2025). These perspectives suggest a question about the *interface between encoding and prediction*: when a model revisits its encoded history, which relations should guide that read operation? Increasing depth alone does not specify the evidence that each additional step should retrieve.

We investigate distance-conditioned refinement. **MeRa** (**Me**tric-space **Rea**soning) retains the encoder output as a fixed memory and repeatedly queries it using an evolving trajectory representation. Pairwise distances enter the attention logits through a learned per-head function, and a sigmoid gate scales each residual update. This arrangement separates three design decisions: what is stored, how it is read, and how the resulting update changes the prediction state. The surrounding encoder and prediction objective can remain unchanged. Here, "reasoning" refers to these finite latent refinement steps; it does not imply symbolic deduction or a demonstrated multi-hop algorithm.

The original results provide a concrete motivation for this interface. Across two backbones and three POI datasets, the reported distance-conditioned variant improves on the corresponding baseline, whereas the reported refinement without distance bias falls below it. On NYC GETNext, the contrast is 0.3969 versus 0.4147 NDCG@10. This is an informative architectural comparison, but its interpretation depends on run identity, aligned geographic inputs, and comparable evaluation. We therefore distinguish the module's mathematical properties from the scope of the historical empirical record.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Table 1: Representative mechanisms and spatial conditioning. This is a design comparison, not a claim of matched accuracy or runtime.

Method	Computational mechanism	Spatial conditioning
GETNext	Graph-enhanced encoding	Trajectory-flow graph
GeoMamba	State-space encoding	Geographic priors
ReaRec / LARES	Latent refinement	General item sequences
ROS	Explicit LLM reasoning	Geographic constraints/rewards
MeRa	Post-encoder refinement	Pairwise-distance attention bias

The **contributions** are:

- **A geometry-conditioned refinement interface.** We formulate post-encoder refinement as repeated access to fixed memory with learned distance-dependent attention offsets and gated updates. The interface accommodates geographic and Euclidean distances without altering the prediction objective.
- **Explicit architectural guarantees and limits.** We establish finite-depth perturbation bounds and characterize access to fixed memory. These results separate update sensitivity from convergence and distinguish repeated retrieval from local message propagation.
- **A structured empirical comparison.** We organize the original results across six POI settings, bias ablations, refinement depths, and a derived geometric task. The analysis identifies the strongest reported contrasts while making protocol and historical implementation uncertainty explicit.

2 Related Work

2.1 Spatial prediction

Spatial recommendation has progressed from spatio-temporal recurrent models (Liu et al., 2016) to attention (Luo et al., 2021; Kang & McAuley, 2018), graph-enhanced Transformers (Yang et al., 2022), hypergraphs (Yan et al., 2023), and state-space models with geographic or hyperbolic structure (Qin et al., 2025; Li et al., 2025). GETNext combines trajectory-flow graph features with user, time, and category representations. These inputs can also be retained when replacing its sequence encoder with an LSTM; an LSTM backbone should therefore not be equated with a predictor lacking graph or user information. Our focus is the placement of an explicit distance-conditioned read operation after sequence encoding.

2.2 Latent reasoning and test-time scaling

ReaRec introduces autoregressive latent reasoning for sequential recommendation (Tang et al., 2026b); LARES studies recurrent-depth reasoning (Liu et al., 2025a), and parallel latent reasoning explores computation across multiple streams (Tang et al., 2026a). Earlier attention-based iterative reasoning includes MAC, while Universal Transformers and recurrent-depth language models examine repeated representation updates (Geiping et al., 2025). These precedents establish iterative computation as a broad design pattern. ROS uses explicit LLM reasoning with geographic constraints and spatial rewards (Lv et al., 2026); it does incorporate geography, but through a different mechanism from continuous attention-logit conditioning. MeRa examines a finite post-encoder refinement stack, rather than claiming that separately trained depths establish test-time scaling of one checkpoint.

2.3 Position and distance biases in attention

Attention can incorporate structural information in its logits or query–key representation. ALiBi and RoPE primarily encode sequence position (Press et al., 2022; Su et al., 2024). More directly relevant, STAN adds spatial and temporal relation information to attention for POI recommendation (Luo et al., 2021). Thus, distance-aware attention itself is established. The question studied here is how to use such conditioning in an added refinement stage that repeatedly reads fixed encoder memory. A learned distance function offers flexibility, but does not impose monotonic distance decay.

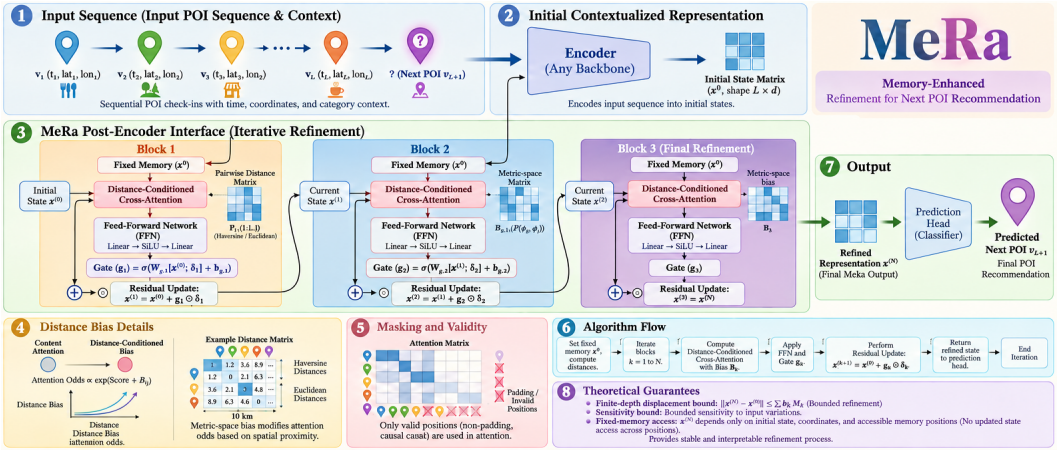


Figure 1: MeRa as a post-encoder interface. Each untied block reads fixed encoder memory. Pairwise distances condition attention, while the gate scales the candidate update before residual addition. Sharing distances does not imply sharing learned bias tensors.

Positioning. Table 1 separates the computational mechanism from its spatial input. This distinction avoids conflating explicit language reasoning, continuous refinement, and spatial encoding.

3 Methodology

3.1 Problem Formulation

Given a user's check-in sequence $\{(v_1, t_1), (v_2, t_2), \ldots, (v_L, t_L)\}$ where each v_i is a POI with geographic coordinates $\mathbf{p}_i = (\text{lat}_i, \text{lon}_i)$ and t_i is the visit time, the task is to predict the next POI v_{L+1}. Each POI v_i is also associated with a category c_i. MeRa is agnostic to the choice of backbone encoder; given any encoder that produces contextualized representations $\mathbf{x}^{(0)} \in \mathbb{R}^{L \times d}$, our module can be inserted before the prediction heads.

3.2 Overview

Figure 1 shows the refinement interface. The initial representation $\mathbf{x}^{(0)}$ serves both as the initial state and as fixed key–value memory. At step k, distance-conditioned cross-attention retrieves information from this memory, an FFN produces an update, and a gate scales its contribution to the state. Distances are shared across steps; learned projections may be step-specific. The code-aligned reference considered here uses untied blocks, including their distance MLPs. This distinction matters for parameter accounting and for interpreting depth.

3.3 Metric-space Bias

Definition 1 (Metric-space bias)**.** *Let (S, ρ) be a metric space and $s_i \in S$ the location associated with position i. A distance-conditioned attention bias for step k and head h is*

$$B_k^{(h)}(i, j) = f_{\theta_k}^{(h)}(-\rho(s_i, s_j)).$$

This quantity is added to the content attention logit before softmax.

We use a two-layer map $\text{Linear}(1, d/4) \to \text{SiLU} \to \text{Linear}(d/4, H)$. The hidden representation is shared across heads within a block; the output projections produce head-specific offsets. Different heads can learn different distance responses, although specialization is not guaranteed. Haversine distance gives the geographic instantiation; Euclidean distance gives the geometric instantiation. The geographic reference uses $\bar{\rho} = \min(\rho, 500\,\text{km})$ as the MLP input distance; D and the bias formulas use this capped distance for that implementation. Locations, rather than categorical POI identifiers, are the points of the metric space: distinct POIs may occupy the same coordinates.

Algorithm 1 MeRa inference

Require: Encoder states $\mathbf{x}^{(0)}$, aligned coordinates $\{\mathbf{p}_i\}$, depth N
Ensure: Refined states $\mathbf{x}^{(N)}$
1: Set fixed memory $\mathbf{M} \leftarrow \mathbf{x}^{(0)}$ and $D_{ij} \leftarrow \rho(\mathbf{p}_i, \mathbf{p}_j)$
2: **for** $k = 0$ to $N - 1$ **do**
3: Compute $\mathbf{B}_k \leftarrow f_{\theta_k}(-D)$
4: $\boldsymbol{\delta}_k \leftarrow \mathrm{FFN}_k(\mathrm{CrossAttn}_k(\mathbf{x}^{(k)}, \mathbf{M}; \mathbf{B}_k))$ with valid-position masks
5: $\mathbf{g}_k \leftarrow \sigma(\mathbf{W}_{g,k}[\mathbf{x}^{(k)}; \boldsymbol{\delta}_k] + \mathbf{b}_{g,k})$
6: $\mathbf{x}^{(k+1)} \leftarrow \mathbf{x}^{(k)} + \mathbf{g}_k \odot \boldsymbol{\delta}_k$
7: **end for**
8: **return** $\mathbf{x}^{(N)}$ to the original prediction heads

Distance decay motivates exposing distances to the model, but an unconstrained MLP does not enforce it. Even with negated distance as input, the learned map need not assign a larger bias to a nearer location. Nor does a larger bias alone guarantee larger final attention, because content logits also contribute. The interpretation of *metric-space* is therefore the source of the conditioning variable, not a claim that the learned bias itself satisfies metric axioms or monotonicity.

For fixed content logits c_{ij} and two unmasked positions j, ℓ, the attention odds satisfy

$$\frac{\alpha_{ij}}{\alpha_{i\ell}} = \exp(c_{ij} - c_{i\ell}) \exp(B_{ij} - B_{i\ell}).$$

This identity isolates the role of the distance channel: it multiplicatively reweights content-based retrieval odds. Unlike a hard spatial neighborhood, a finite bias does not remove an otherwise admissible memory position.

3.4 Reasoning Module

Let $\mathbf{M} = \mathbf{x}^{(0)} \in \mathbb{R}^{L \times d}$ denote fixed memory. The kth block applies

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{g}_k \odot \boldsymbol{\delta}_k, \qquad \boldsymbol{\delta}_k = \mathrm{FFN}_k(\mathrm{CrossAttn}_k(\mathbf{x}^{(k)}, \mathbf{M}; \mathbf{B}_k)), \tag{1}$$

with $\mathbf{g}_k = \sigma(\mathbf{W}_{g,k}[\mathbf{x}^{(k)}; \boldsymbol{\delta}_k] + \mathbf{b}_{g,k})$. Linear maps act row-wise; gates have the same $L \times d$ shape as the state. In each attention head,

$$\mathrm{Attn}(Q, K, V) = \mathrm{softmax}\Big(QK^\top / \sqrt{d_k} + \mathbf{B}_k + \mathbf{M}_{\mathrm{causal}} + \mathbf{M}_{\mathrm{pad}}\Big) V, \tag{2}$$

where queries depend on the current state, keys and values are projections of $\mathbf{M}$, and $d_k = d/H$. The notation suppresses the block index on the projections. Masked positions are excluded from valid reads in this intended operation; Appendix E distinguishes this specification from unresolved masking paths in the archived source.

The reference block normalizes the query and fixed memory separately, then applies a normalized $d \to 4d \to d$ feed-forward path with GELU and dropout. The residual gate regulates the amplitude of each update; it does not certify that an update improves accuracy. Setting the update to zero recovers the identity map. Small gates likewise permit small modifications, without guaranteeing that optimization will choose them. Fixed memory preserves the information source across steps, but does not by itself prevent cumulative state drift.

3.5 Theoretical Results

We characterize finite-depth updates and information access, rather than asserting convergence or a strict expressivity hierarchy. Throughout this analysis, the encoder memory is fixed, parameters may differ across steps, and inference is deterministic (dropout is disabled). Norms are Euclidean after flattening the sequence representations.

Theorem 1 (Finite-depth displacement and sensitivity)**.** *For fixed memory, write step k as $F_k(\mathbf{x}) = \mathbf{x} + \mathbf{g}_k(\mathbf{x}) \odot \boldsymbol{\delta}_k(\mathbf{x})$. Suppose, on a set containing the trajectories under consideration, $\|\mathbf{g}_k(\mathbf{x})\|_\infty \leq$*

Table 2: Originally reported dataset statistics. CLEVR entries distinguish objects per scene from the scene count; historical split identities remain unresolved.

Statistic	NYC	TKY	CA	CLEVR
# POIs / Objects	4,980	7,832	9,689	3–10 (per scene)
# Users / Scenes	1,083	2,293	3,112	15,000
# Check-ins / Scenes	48,547	119,834	153,462	15,000
# Categories	251	328	387	24
Avg. trajectory length	7.8	8.2	6.9	6.5
Distance metric	Haversine	Haversine	Haversine	Euclidean

$b_k \le 1$ *and* $\|\boldsymbol{\delta}_k(\mathbf{x})\| \le M_k$. *Then*

$$\|\mathbf{x}^{(N)} - \mathbf{x}^{(0)}\| \le \sum_{k=0}^{N-1} b_k M_k. \tag{3}$$

If, additionally, $\boldsymbol{\delta}_k$ *and* $\mathbf{g}_k$ *are Lipschitz on that set with constants* L_k *and* G_k*, respectively, two trajectories using the same memory and parameters satisfy*

$$\|\mathbf{x}^{(N)} - \widetilde{\mathbf{x}}^{(N)}\| \le \left[\prod_{k=0}^{N-1} (1 + b_k L_k + M_k G_k)\right] \|\mathbf{x}^{(0)} - \widetilde{\mathbf{x}}^{(0)}\|. \tag{4}$$

The second statement concerns perturbations of the initial query state while holding memory fixed; it is not an end-to-end encoder robustness bound. Smaller gate magnitudes limit the direct update term, but gate sensitivity also contributes. These bounds neither establish contraction nor guarantee improved predictions. Proofs are in Appendix B.

Theorem 2 (Fixed-memory access)**.** *Let* $\mathbf{m}_j$ *and* $\mathbf{p}_j$ *be the fixed encoder memory and coordinate at position* j*. For each query position* i*, let* A_i *be a fixed, nonempty set of allowed memory positions. Suppose every step uses keys and values derived positionwise from* $\{\mathbf{m}_j : j \in A_i\}$*, with bias* $f_{\theta_k}(-\rho(\mathbf{p}_i, \mathbf{p}_j))$*, and that query projections, normalization, FFNs, and gates act positionwise. Then, at every finite depth* N*,* $\mathbf{x}_i^{(N)}$ *is a function only of*

$$\mathbf{x}_i^{(0)}, \quad \mathbf{p}_i, \quad \{(\mathbf{m}_j, \mathbf{p}_j) : j \in A_i\}, \tag{5}$$

and the fixed model parameters and mask. In particular, refinement does not acquire another position's updated state by reading its memory.

For the causal attention used here, A_i comprises the allowed nonpadding history positions. A finite distance bias does not impose a hard spatial radius. The statement also applies if a fixed local mask is introduced, but such a mask would not turn fixed memory into dynamic message passing. Encoder outputs may already contain contextual information from other positions. The theorem therefore characterizes access to encoder memory, not locality with respect to raw inputs, and it makes no claim that different depths realize identical or strictly nested function classes.

4 Experiments

4.1 Settings

Datasets. The original study reports three POI datasets following a GETNext-style preprocessing pipeline (Yang et al., 2022): Foursquare NYC and TKY, and Gowalla CA. Its stated protocol removes trajectories shorter than two check-ins and uses a chronological 80/20 training–validation split. Table 2 preserves the reported POI statistics; the current archived node tables differ in TKY and CA size, so these statistics are not independently certified historical inputs. The CLEVR-derived setting uses scene annotations (Johnson et al., 2017) and a Euclidean-nearest-object target.

Backbones and baselines. We retain the original comparisons for GETNext and an LSTM sequence backbone. GETNext uses graph-derived POI representations and user, time, and category inputs (Yang

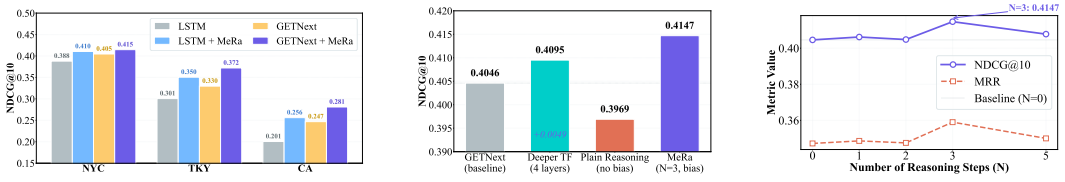
Table 3: Preserved original ranking scores. Upper block: external literature values compiled in Li et al. (2025), whose stated split is 8:1:1. Lower blocks: original backbone comparisons under the stated 80/20 validation protocol. Cross-block ranks are not treated as matched performance comparisons. *Originally marked as reproduced; all historical run identities remain to be bound.

Method	NYC				TKY				CA			
	ND@1	ND@5	ND@10	MRR	ND@1	ND@5	ND@10	MRR	ND@1	ND@5	ND@10	MRR
PLSPL	0.1601	0.3048	0.3336	0.2849	0.1495	0.2831	0.3143	0.2642	0.1084	0.1759	0.2029	0.1678
HME	0.1619	0.2806	0.3226	0.2787	0.1535	0.2637	0.2924	0.2366	0.1181	0.1886	0.2232	0.1945
AGRAN	0.2202	0.3638	0.3792	0.3343	0.1755	0.2989	0.3261	0.2879	0.1329	0.2121	0.2331	0.2043
MCLP	0.2404	0.3674	0.3973	0.3507	0.1662	0.3110	0.3415	0.3199	0.1324	0.1914	0.2121	0.1895
GeoMamba'24	0.1988	0.3392	0.3506	0.3246	0.1851	0.2953	0.3205	0.2858	0.1256	0.2029	0.2215	0.1962
GeoMamba'25	0.2377	0.3786	0.4012	0.3566	0.2157	0.3402	0.3686	0.3209	0.1388	0.2485	0.2754	0.2373
$\text{HMamba}_{\text{full}}$	0.2204	0.3679	0.4031	0.3465	0.1828	0.3341	0.3673	0.3127	0.1366	0.2501	0.2792	0.2421
$\text{HMamba}_{\text{half}}$	0.1896	0.3453	0.3767	0.3222	0.1945	0.3295	0.3603	0.3118	0.1423	0.2381	0.2648	0.2317
HMST	0.2138	0.3747	0.4063	0.3482	0.1925	0.3325	0.3690	0.3257	0.1356	0.2325	0.2680	0.2300
HVGAE	0.2271	0.3651	0.3982	0.3470	0.1977	0.3167	0.3455	0.3180	0.1391	0.2325	0.2658	0.2367
LSTM*	0.2320	0.3580	0.3880	0.3420	0.1715	0.2761	0.3006	0.2637	0.1143	0.1819	0.2005	0.1775
LSTM + MeRa	0.2395	0.3790	0.4105	0.3565	0.1942	0.3198	0.3505	0.3052	0.1460	0.2310	0.2561	0.2247
improv.	*↑3.2%*	*↑5.9%*	*↑5.8%*	*↑4.2%*	*↑13.2%*	*↑15.8%*	*↑16.6%*	*↑15.7%*	*↑27.7%*	*↑27.0%*	*↑27.7%*	*↑26.6%*
GETNext	0.2244	0.3736	0.4046	0.3472	0.1767	0.3072	0.3297	0.2934	0.1342	0.2188	0.2468	0.2121
GETNext + MeRa	0.2425	0.3825	0.4147	0.3590	0.1980	0.3428	0.3720	0.3268	0.1562	0.2561	0.2810	0.2459
improv.	*↑8.1%*	*↑2.4%*	*↑2.5%*	*↑3.4%*	*↑12.1%*	*↑11.6%*	*↑12.8%*	*↑11.4%*	*↑16.4%*	*↑17.0%*	*↑13.9%*	*↑15.9%*

et al., 2022). The archived LSTM path retains such contextual inputs; the backbone label does not establish a spatially naive system. Literature rows cover PLSPL, HME, AGRAN, MCLP, GeoMamba, HMamba, HMST, and HVGAE (Wu et al., 2022; Feng et al., 2020; Wang et al., 2023; Sun et al., 2024; Qin et al., 2025; Zhang et al., 2026; Qiao et al., 2025; Liu et al., 2025b). These rows are taken from the original comparison source (Li et al., 2025) and provide context rather than a harmonized leaderboard.

Metrics. We report NDCG@1, NDCG@5, NDCG@10, and MRR, computed by ranking over the candidate vocabulary. The study's own results are reported under the stated validation protocol. The implementation details are in Appendix E.

Implementation details. The default reported depth is $N = 3$ with two attention heads. The available source combines 128-dimensional POI and user features with 32-dimensional time and category features, giving refinement width $d = 320$ and bias hidden width 80. Its blocks are untied and use dropout 0.3. These reference specifications differ from the original reported $d = 128$, bias width 32, and module dropout 0.1; Appendix C records both. The original hardware and optimizer record is retained separately from code verification. Without bound historical configurations, neither specification can retrospectively certify the original table rows.



(a) Backbone comparison. (b) Ablation study. (c) Reasoning depth.

Figure 2: Original reported comparisons. (a) Reported NDCG@10 for each backbone pair. (b) NYC GETNext baseline, deeper Transformer, and refinement variants. (c) Reported NYC GETNext NDCG@10 and MRR by depth. Original plots are retained; the bar charts in (a,b) use truncated vertical axes, so bar-height ratios do not represent metric ratios.

4.2 Comparison

Table 3 is most informative when read through the within-backbone contrasts. GETNext NDCG@10 increases from 0.4046 to 0.4147 on NYC, 0.3297 to 0.3720 on TKY, and 0.2468 to 0.2810 on CA,

Table 4: Original bias-ablation scores. In all six reported settings, the variant without bias is below its baseline and MeRa is above it. This describes the record, not a universal necessity result.

Dataset	Configuration	ND@1	ND@5	ND@10	MRR
NYC	LSTM (baseline)	0.2320	0.3580	0.3880	0.3420
	+ Reasoning (w/o bias)	0.2285	0.3526	0.3822	0.3370
	+ MeRa (N=3)	**0.2395**	**0.3790**	**0.4105**	**0.3565**
	GETNext (baseline)	0.2244	0.3736	0.4046	0.3472
	+ Reasoning (w/o bias)	0.2201	0.3665	0.3969	0.3406
	+ MeRa (N=3)	**0.2425**	**0.3825**	**0.4147**	**0.3590**
TKY	LSTM (baseline)	0.1715	0.2761	0.3006	0.2637
	+ Reasoning (w/o bias)	0.1681	0.2706	0.2946	0.2584
	+ MeRa (N=3)	**0.1942**	**0.3198**	**0.3505**	**0.3052**
	GETNext (baseline)	0.1767	0.3072	0.3297	0.2934
	+ Reasoning (w/o bias)	0.1740	0.3026	0.3247	0.2890
	+ MeRa (N=3)	**0.1980**	**0.3428**	**0.3720**	**0.3268**
CA	LSTM (baseline)	0.1143	0.1819	0.2005	0.1775
	+ Reasoning (w/o bias)	0.1114	0.1774	0.1955	0.1731
	+ MeRa (N=3)	**0.1460**	**0.2310**	**0.2561**	**0.2247**
	GETNext (baseline)	0.1342	0.2188	0.2468	0.2121
	+ Reasoning (w/o bias)	0.1315	0.2144	0.2419	0.2079
	+ MeRa (N=3)	**0.1562**	**0.2561**	**0.2810**	**0.2459**

corresponding to relative improvements of 2.5%, 12.8%, and 13.9%. The LSTM pairs show gains of 5.8%, 16.6%, and 27.7%, respectively. The positive direction is shared by all six reported pairs, while its magnitude varies substantially.

The contribution of this comparison is the repeated contrast across two sequence encoders, not a cross-paper state-of-the-art claim. The external rows have not been re-evaluated under the study's split. The source paper's own GTR-Mamba results are also outside a matched comparison; the selected literature rows are not exhaustive. Consequently, differences from individual external methods cannot isolate an architectural advantage. Even the within-study contrasts remain conditional on historical input and run alignment, discussed in Section F.

4.3 Ablation Study

The ablation in Table 4 asks whether additional refinement benefits from an explicit distance channel. Three aspects of the original record deserve separate interpretation.

Distance conditioning. On NYC GETNext, refinement without bias gives 0.3969 NDCG@10, versus a baseline of 0.4046 and a distance-conditioned value of 0.4147. The latter exceeds the no-bias variant by 0.0178, or 4.5% relative to that variant. The same ordering appears in all six reported settings. This pattern motivates distance conditioning in this implementation family, but does not establish that every successful spatial refinement method requires it. Single-seed point estimates cannot establish run-to-run reliability.

Capacity. The original four-layer Transformer control reports 0.4095 on NYC, below the three-step MeRa result of 0.4147. This comparison indicates that the reported benefit is not reproduced by that particular depth expansion. It does not eliminate capacity or optimization as explanations: parameter accounting and historical configurations are unresolved, and equal training limits do not imply equal realized updates.

Gating. The gate scales residual updates in the specified architecture. A comparison that disables only the bias does not isolate the gate's contribution. We therefore do not attribute the bias contrast to protection from harmful updates. A causal interpretation further requires that the original runs differ only in the intended intervention; that identity has not been independently reconstructed.

Table 5: Effect of reasoning depth N (NDCG@10).

Backbone	Dataset	N=0	N=1	N=2	N=3	N=5
LSTM	NYC	0.3880	0.3935	0.4015	**0.4105**	0.4028
	TKY	0.3006	0.3148	0.3325	**0.3505**	0.3390
	CA	0.2005	0.2155	0.2358	**0.2561**	0.2450
GETNext	NYC	0.4046	0.4062	0.4048	**0.4147**	0.4078
	TKY	0.3297	0.3405	0.3558	**0.3720**	0.3598
	CA	0.2468	0.2580	0.2695	**0.2810**	0.2738

Table 6: Original CLEVR-derived nearest-object scores. Task identity and candidate encoding require historical verification; these are not standard CLEVR VQA results.

Method	NDCG@1	NDCG@5	NDCG@10	MRR
Seq Transformer	0.1323	0.4109	0.5093	0.3580
+ Reasoning (w/o bias)	0.1284	0.4021	0.5038	0.3497
+ MeRa (Euclidean bias)	**0.1504**	**0.4312**	**0.5319**	**0.3796**

4.4 Effect of Reasoning Depth

Depth changes the number of opportunities to query memory, but its empirical value need not increase monotonically. On NYC GETNext, the reported NDCG@10 values are 0.4046, 0.4062, 0.4048, 0.4147, and 0.4078 for $N = 0, 1, 2, 3, 5$. Across all six settings in Table 5, the largest reported value occurs at three steps and decreases at five.

These observations support three steps as the original study's selected operating point. They do not establish a universal optimal depth or convergence of a learned state. In an untied stack, depth also changes parameter count and optimization. The historical records do not identify this table as evaluation of one fixed checkpoint at varying inference depths. We therefore interpret it as a reported depth comparison, without claiming test-time scaling or assigning the five-step decline to an unmeasured optimization mechanism.

4.5 Beyond Geographic Distance: CLEVR Spatial Reasoning

The original CLEVR-derived task replaces geographic distance with Euclidean distance on scene annotations (Johnson et al., 2017). Its stated target is the nearest remaining object to the last observed object in a randomized sequence. Table 6 preserves the original scores: NDCG@10 is 0.5093 for the Transformer, 0.5038 with refinement without bias, and 0.5319 with the Euclidean-conditioned variant, a reported 4.4% increase over the Transformer.

This record is suggestive of a second distance-based prediction setting, but it does not establish general geometric or multi-hop reasoning. Nearest-object prediction depends critically on which candidate coordinates and identities are visible. The available converter also raises a target-ID/coordinate alignment concern, and its exact connection to these runs is unresolved. Appendix D states the information requirements and provenance limits. We do not interpret the table as a verified transfer experiment, nor compare it to standard CLEVR visual question answering.

4.6 Additional Analysis

Why plain reasoning hurts. The reported decline of the no-bias variant raises a mechanism question rather than answering it. Additional content-based reads may change useful encoder states without supplying a useful spatial relation. Distance conditioning offers one way to change those reads, as the attention-odds identity shows. However, content attention is not uniform, and the observed scores do not establish that it ignores geography or dilutes a specific signal. Attention analysis and matched training evidence would be needed to distinguish these explanations.

Table 7: Original NYC cost record. Parameter counts and timings are historical reports, not verified measurements of the code-aligned reference.

Configuration	Reported Params	Time/Epoch	Overhead
GETNext (baseline)	0	142s	–
+ MeRa (N=1)	0.40M	147s	+3.5%
+ MeRa (N=3)	1.20M	153s	+7.7%
+ MeRa (N=5)	2.00M	161s	+13.4%
LSTM (baseline)	0	58s	–
+ MeRa (N=1)	0.40M	62s	+6.9%
+ MeRa (N=3)	1.20M	68s	+17.2%
+ MeRa (N=5)	2.00M	75s	+29.3%

Backbone-dependent gains. The larger relative gains on the LSTM backbone are consistent with the value of a refinement interface depending on the starting representation. They do not show that the LSTM receives its first spatial information from MeRa: the archived input path includes graph-derived POI and user features. Differences in encoder capacity, baseline quality, and optimization may all contribute. The modular interface is useful precisely because it exposes the same design question to different encoders without requiring their inputs to be identical in expressive power.

Complementarity principle. A working hypothesis is that refinement is most useful when its read operation supplies relations not already exposed effectively to the head. The present results motivate this hypothesis, rather than establishing a law of diminishing returns. The practical lesson is to assess the *conditioning of additional computation* alongside its amount. Distance bias, memory choice, and residual updates form distinct design axes whose isolated contributions remain open.

Computational overhead. For fixed width, each refinement step adds attention with quadratic dependence on sequence length and an FFN with linear dependence on sequence length. Table 7 retains the original timing record: at three steps, reported training overhead is 7.7% for GETNext and 17.2% for LSTM. These are not independently reproduced measurements. The reported 1.20M added parameters conflict with the available $d = 320$ reference, whose three blocks contain 4,317,126 parameters by static accounting. The table therefore cannot certify that reference's parameter efficiency. Distances can be reused, but learned biases depend on block parameters; no inference-latency advantage is claimed without measurements.

5 Discussion

The six reported POI comparisons motivate distance-conditioned post-encoder refinement as a useful design direction. However, single-seed validation results and unresolved historical run alignment limit causal interpretation and generalization. Our theoretical results characterize the refinement interface, rather than guaranteeing convergence or predictive improvement. Appendix **??** details these limitations, additional evidence, and broader impacts.

6 Conclusion

MeRa treats post-encoder computation as a distance-conditioned read–update process. Fixed memory separates stored evidence from evolving queries, and gated residuals control update magnitude. The original record reports favorable comparisons across two backbones and three POI datasets, while depth results show that more computation is not automatically better. Our analysis makes the interface's guarantees precise and separates them from claims the evidence cannot sustain. The interface makes memory, distance conditioning, and update magnitude explicit design choices. The reported contrasts motivate testing their roles under fully traceable input and run configurations.

AI USE STATEMENT

AI tools assisted with language checking and polishing. No additional model-training or inference experiments were performed as part of these revisions. The authors take responsibility for the accuracy of the reported results, citations, and final manuscript.

ETHICS STATEMENT

This work studies distance-conditioned refinement for next-location prediction using existing POI datasets. Location histories can reveal sensitive information about individuals, and the proposed method provides no privacy guarantee. The reported results are limited to the evaluated settings and do not establish reliability, fairness, or suitability for deployment. Any practical use should ensure appropriate consent and data protection, assess performance disparities across users and geographic regions, and avoid unauthorized tracking or profiling.

REPRODUCIBILITY STATEMENT

Code is available at https://anonymous.4open.science/r/MeRaa. The repository includes a reference training configuration, saved prediction records, and scripts for recomputing statistical summaries. Experimental settings and evaluation protocols are documented in the main text and appendix.

REFERENCES

Shanshan Feng, Lucas Vinh Tran, Gao Cong, Lisi Chen, Jing Li, and Fan Li. HME: A hyperbolic metric embedding approach for next-POI recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1429–1438, 2020. doi: 10.1145/3397271.3401049. URL https://doi.org/10.1145/3397271.3401049.

Jonas Geiping, Sean McLeish, Neel Jain, John Kirchenbauer, Siddharth Singh, Brian R. Bartoldson, Bhavya Kailkhura, Abhinav Bhatele, and Tom Goldstein. Scaling up test-time compute with latent reasoning: A recurrent depth approach. In *Advances in Neural Information Processing Systems*, volume 38, pp. 46242–46293, 2025. doi: 10.52202/085713-1380. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/3b01972cf31e6fa0fe29e4b8b5c2a0a1-Abstract-Conference.html. Spotlight.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2901–2910, 2017. URL https://openaccess.thecvf.com/content_cvpr_2017/html/Johnson_CLEVR_A_Diagnostic_CVPR_2017_paper.html.

Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *2018 IEEE International Conference on Data Mining (ICDM)*, pp. 197–206, 2018. doi: 10.1109/ICDM.2018.00035. URL https://doi.org/10.1109/ICDM.2018.00035.

Zhuoxuan Li, Jieyuan Pei, Tangwei Ye, Zhongyuan Lai, Zihan Liu, Fengyuan Xu, Qi Zhang, and Liang Hu. GTR-Mamba: Geometry-to-tangent routing Mamba for hyperbolic POI recommendation. *arXiv preprint arXiv:2510.22942*, 2025. doi: 10.48550/arXiv.2510.22942. URL https://arxiv.org/abs/2510.22942.

Enze Liu, Bowen Zheng, Xiaolei Wang, Wayne Xin Zhao, Jinpeng Wang, Sheng Chen, and Ji-Rong Wen. LARES: Latent reasoning for sequential recommendation. *arXiv preprint arXiv:2505.16865*, 2025a. doi: 10.48550/arXiv.2505.16865. URL https://arxiv.org/abs/2505.16865.

Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, pp. 194–200, 2016. doi: 10.1609/aaai.v30i1.9971. URL https://ojs.aaai.org/index.php/AAAI/article/view/9971.

Yuwen Liu, Lianyong Qi, Xingyuan Mao, Weiming Liu, Fan Wang, Xiaolong Xu, Xuyun Zhang, Wanchun Dou, Xiaokang Zhou, and Amin Beheshti. Hyperbolic variational graph auto-encoder for next POI recommendation. In *Proceedings of the ACM on Web Conference 2025*, pp. 3267–3275, 2025b. doi: 10.1145/3696410.3714804. URL https://doi.org/10.1145/3696410.3714804.

Yingtao Luo, Qiang Liu, and Zhaocheng Liu. STAN: Spatio-temporal attention network for next location recommendation. In *Proceedings of the Web Conference 2021*, pp. 2177–2185, 2021. doi: 10.1145/3442381.3449998. URL https://doi.org/10.1145/3442381.3449998.

Dongyi Lv, Qiuyu Ding, Heng-Da Xu, Zhaoxu Sun, Zhi Wang, Feng Xiong, and Mu Xu. Reasoning over space: Enabling geographic reasoning for LLM-based generative next POI recommendation. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 7322–7336. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.332. URL https://aclanthology.org/2026.acl-long.332/.

Ofir Press, Noah A. Smith, and Mike Lewis. Train short, test long: Attention with linear biases enables input length extrapolation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=R8sQPpGCv0.

Hongliang Qiao, Shanshan Feng, Min Zhou, WenTao Li, and Fan Li. Hyperbolic multi-semantic transition for next POI recommendation. In *Companion Proceedings of the ACM on Web Conference 2025*, pp. 1830–1837, 2025. doi: 10.1145/3701716.3717802. URL https://doi.org/10.1145/3701716.3717802.

Yifang Qin, Jiaxuan Xie, Zhiping Xiao, and Ming Zhang. GeoMamba: Towards multi-granular POI recommendation with geographical state space model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 12479–12487, 2025. doi: 10.1609/aaai.v39i12.33360. URL https://ojs.aaai.org/index.php/AAAI/article/view/33360.

Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063. URL https://doi.org/10.1016/j.neucom.2023.127063.

Tianao Sun, Ke Fu, Weiming Huang, Kai Zhao, Yongshun Gong, and Meng Chen. Going where, by whom, and at what time: Next location prediction considering user preference and temporal regularity. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2784–2793, 2024. doi: 10.1145/3637528.3671916. URL https://doi.org/10.1145/3637528.3671916.

Jiakai Tang, Xu Chen, Wen Chen, Jian Wu, Yuning Jiang, and Bo Zheng. Parallel latent reasoning for sequential recommendation. *arXiv preprint arXiv:2601.03153*, 2026a. doi: 10.48550/arXiv.2601.03153. URL https://arxiv.org/abs/2601.03153.

Jiakai Tang, Sunhao Dai, Teng Shi, Jun Xu, Xu Chen, Wen Chen, Jian Wu, and Yuning Jiang. Think before recommend: Unleashing the latent reasoning power for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 38(9):6226–6239, 2026b. doi: 10.1109/TKDE.2026.3694421. URL https://doi.org/10.1109/TKDE.2026.3694421.

Zhaobo Wang, Yanmin Zhu, Chunyang Wang, Wenze Ma, Bo Li, and Jiadi Yu. Adaptive graph representation learning for next POI recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 393–402, 2023. doi: 10.1145/3539618.3591634. URL https://doi.org/10.1145/3539618.3591634.

Yuxia Wu, Ke Li, Guoshuai Zhao, and Xueming Qian. Personalized long- and short-term preference learning for next POI recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 34(4):1944–1957, 2022. doi: 10.1109/TKDE.2020.3002531. URL https://doi.org/10.1109/TKDE.2020.3002531.

Xiaodong Yan, Tengwei Song, Yifeng Jiao, Jianshan He, Jiaotuan Wang, Ruopeng Li, and Wei Chu. Spatio-temporal hypergraph learning for next POI recommendation. In *Proceedings of the 46th*

International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 403–412, 2023. doi: 10.1145/3539618.3591770. URL https://doi.org/10.1145/3539618.3591770.

Song Yang, Jiamou Liu, and Kaiqi Zhao. GETNext: Trajectory flow map enhanced transformer for next POI recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1144–1153, 2022. doi: 10.1145/3477495.3531983. URL https://doi.org/10.1145/3477495.3531983.

Qianru Zhang, Honggang Wen, Wei Yuan, Crystal Chen, Menglin Yang, Siu-Ming Yiu, and Hongzhi Yin. HMamba: Hyperbolic Mamba for sequential recommendation. *ACM Transactions on Information Systems*, 44(5):106:1–106:27, 2026. doi: 10.1145/3811405. URL https://doi.org/10.1145/3811405.

A Appendix

B Detailed Proofs

B.1 Proof of Theorem 1

For any vectors $\mathbf{a}, \mathbf{b}$, $\|\mathbf{a} \odot \mathbf{b}\| \le \|\mathbf{a}\|_\infty \|\mathbf{b}\|$. Applying this inequality to each residual update and then telescoping gives

$$\|\mathbf{x}^{(N)} - \mathbf{x}^{(0)}\| \le \sum_{k=0}^{N-1} \|\mathbf{g}_k(\mathbf{x}^{(k)}) \odot \boldsymbol{\delta}_k(\mathbf{x}^{(k)})\| \tag{6}$$

$$\le \sum_{k=0}^{N-1} b_k M_k. \tag{7}$$

For the sensitivity statement, adding and subtracting $\mathbf{g}_k(\mathbf{x}) \odot \boldsymbol{\delta}_k(\mathbf{y})$ yields

$$\begin{aligned} F_k(\mathbf{x}) - F_k(\mathbf{y}) = \mathbf{x} - \mathbf{y} &+ \mathbf{g}_k(\mathbf{x}) \odot (\boldsymbol{\delta}_k(\mathbf{x}) - \boldsymbol{\delta}_k(\mathbf{y})) \\ &+ (\mathbf{g}_k(\mathbf{x}) - \mathbf{g}_k(\mathbf{y})) \odot \boldsymbol{\delta}_k(\mathbf{y}). \end{aligned} \tag{8}$$

The first product has norm at most $b_k L_k \|\mathbf{x} - \mathbf{y}\|$. For the second, use $\|\mathbf{g}_k(\mathbf{x}) - \mathbf{g}_k(\mathbf{y})\|_\infty \le \|\mathbf{g}_k(\mathbf{x}) - \mathbf{g}_k(\mathbf{y})\| \le G_k \|\mathbf{x} - \mathbf{y}\|$. Thus

$$\|F_k(\mathbf{x}) - F_k(\mathbf{y})\| \le (1 + b_k L_k + M_k G_k) \|\mathbf{x} - \mathbf{y}\|. \tag{9}$$

Applying this bound successively proves Eq. equation 4. The factor is not required to be less than one; no fixed-point conclusion follows. □

B.2 Proof of Theorem 2

Fix a position i and the parameters and mask. Let Z_i denote the inputs listed in Eq. equation 5. At depth zero, $\mathbf{x}_i^{(0)}$ is a component of Z_i. Suppose $\mathbf{x}_i^{(k)}$ is a function of Z_i. Its query projection is then a function of Z_i. Every allowed key, value, and distance bias is also a function of Z_i, so the attention logits, their softmax over A_i, and the weighted value sum depend only on Z_i. Positionwise normalization, the FFN, the gate, and the residual update preserve this dependency. Hence $\mathbf{x}_i^{(k+1)}$ is a function of Z_i. Induction establishes the result for every finite N, including when parameters and bias functions differ across steps.

This argument concerns the refinement module conditional on its encoder memory. It does not remove dependencies already encoded in $\mathbf{m}_j$, and it does not assert that every allowed input has a nonzero influence for every parameter setting. □

C Full Hyperparameter Configuration

Table A1 separates originally reported settings from the available source. This is a configuration comparison, not an assertion that the available source generated every original score. The component embedding width of 128 is not the concatenated refinement width. In the reference, POI/user/time/category widths sum to 320.

Table A1: Reported configuration and code-aligned reference. A shared entry records the original description, not a recovered run identity.

Setting	Original report	Available reference
Refinement width	128	320
Bias hidden width	32	80
Module dropout	0.1	0.3
Attention heads	2	2
Default depth	3	3
Step parameters	Not clearly specified	Untied
Bias parameters	Shared in description	Block-specific
Optimizer / learning rate	Adam / 10^{-3}	Original setting retained
Batch size / weight decay	20 / 5×10^{-4}	Original setting retained
Maximum epochs / patience	200 / 40	Run histories require binding
Random seed	42	Actual initialization not certified
Hardware	RTX 4090	Original timing report

For the available reference block, including its attention projections, feed-forward layers, gate, three LayerNorms, and bias MLP, static accounting gives $14d^2+16d+4\lfloor d/4 \rfloor+2$ parameters. At $d = 320$, this is 1,439,042 per block and 4,317,126 for three blocks. At $d = 128$, the corresponding per-block count is 231,554. Neither recovers the original nominal 0.40M per step. Timing measurements cannot be transferred between these configurations without a bound implementation and measurement procedure.

D CLEVR Dataset Construction

Task definition. The original description uses CLEVR scenes with 3–10 objects, randomizes object order, and predicts the nearest remaining object in the (x, y) plane. Formally, for an observed prefix ending at o_i and remaining candidate set C_i, the intended target is $y_i = \arg\min_{o \in C_i} \|p(o) - p(o_i)\|_2$, with a deterministic rule required for ties. This equation specifies target construction, not a claim that every candidate feature is available to the predictor. If candidate coordinates are available, the target can be computed directly; if they are withheld, the observable inputs and candidate identity scheme must explain how the target is inferable.

Object encoding. The original record describes 24 color–shape category IDs and coordinate-derived spatial features. It does not provide a fully bound description of scene-specific object IDs, candidate masks, or coordinate visibility for the reported runs. A category identifier alone does not distinguish multiple objects sharing that category. The available converter can assign the nearest object's attributes while retaining another remaining object's ID, raising a conditional target-alignment concern. This observation is not an estimated error rate for the historical data.

Split. The original description reports 10,500 training scenes and 4,500 validation scenes drawn from the corresponding CLEVR partitions. These scene counts do not by themselves specify the number of prefix prediction examples. The historical split and converter identities have not been recovered. Consequently, the retained scores document the original report; they do not validate unseen-scene object identity handling, general spatial reasoning, or transfer from the POI models.

E Implementation Details

Haversine distance computation. For aligned coordinates, the intended distance is $\rho = 2R \arcsin(\sqrt{\min(1, \max(0, a))})$, where $a = \sin^2(\Delta\phi/2) + \cos\phi_1 \cos\phi_2 \sin^2(\Delta\lambda/2)$ and $R = 6371$ km. The reference clips large distances at 500 km. A distance matrix can be reused across blocks, whereas untied learned bias tensors are computed with each block's parameters.

Padding and masking. Only valid trajectory positions should contribute to attention and loss. All-masked softmax rows require explicit handling; multiplying a NaN by zero is not a general remedy. The available source raises unresolved batch/sequence-axis, padding, and filtered-timestamp alignment concerns. We therefore specify the intended masked operation in the method without certifying every historical execution path.

Multi-task loss. The original training framework combines POI and category cross-entropy with time-prediction MSE: $\mathcal{L} = \mathcal{L}_{\text{poi}} + \mathcal{L}_{\text{cat}} + \lambda_t \mathcal{L}_{\text{time}}$, with reported $\lambda_t = 10$. The refinement interface leaves the prediction heads and this objective in place.

Reproducibility. A valid implementation must align the POI identifier, graph-feature row, coordinate row, and output class. The available loader extracts coordinates in node-file order but constructs its identifier map from a set. Read-only reconstruction on current node tables found coordinate mismatches for 4,979–4,980 of 4,980 NYC rows, 7,791–7,793 of 7,793 TKY rows, and 9,755 of 9,756 CA rows across three tested hash seeds. These are current-file checks, not historical-run error rates. The original mappings and input hashes are unavailable, so their effect on the original scores remains unknown. Merely fixing the present loader would not certify the old results.

Located validation logs also do not reproduce every original table row under a best-NDCG@10 selection rule. The current revision preserves the table values rather than substituting those log maxima. The separately available later implementation reports mean test NDCG@10 of 0.3792/0.3730, 0.3679/0.3628, and 0.1685/0.1684 for encoder/three-step geometry on NYC/TKY/CA, respectively. These results use a different protocol, provide no general positive replication of the original pattern, and are not pooled with its validation values. All new work in this revision is document preparation, source analysis, or deterministic arithmetic; no new training, validation prediction, or test evaluation is performed.

F Extend Discussion

Limitations. The strongest evidence in the original record is the shared direction of the six POI contrasts. Its principal limitation is historical provenance: the table rows have not been bound to exact code, split files, checkpoint choices, or POI mappings. The available loader constructs POI indices from a set while retaining node-file coordinate order, which can misalign geometry; current-file checks demonstrate this risk, but do not establish its effect on the historical runs. Correct input alignment is a prerequisite for a causal claim about spatial conditioning, not a detail that favorable scores can resolve.

The reported single-seed validation results do not establish statistical significance, held-out test performance, or universal necessity. Separate later experiments with a different implementation and protocol did not show a general advantage for three-step geographic refinement; they are not pooled with the original results or presented as a direct reproduction. The theoretical properties describe the interface, not a convergence or accuracy guarantee. The CLEVR-derived and cost records remain supplementary observations with unresolved implementation identity. These limits leave a focused contribution: an explicit refinement interface and a documented empirical pattern that warrants evaluation under fully traceable conditions.

Broader impact. More accurate location prediction can support recommendation and mobility services, but can also expose sensitive behavioral patterns. The module supplies no privacy or fairness guarantee. Deployment requires appropriate consent, data protection, and evaluation across users and geographic regions; aggregate ranking scores alone cannot establish suitability.