

Destination Support Restoration for Finite-Set Multimodal Trajectory Prediction

Fengrui Liu¹, Jiajun Peng², Duo Peng³, and Feng Liu^{4,*}

Abstract—Robots operating around pedestrians often reason over a finite set of predicted human futures. Repeated online updates can concentrate this limited prediction budget on dominant destinations and leave plausible alternatives under-represented or absent, removing those alternatives from the finite representation available to downstream decision making. We introduce Destination Support Restoration (DSR), a causal post-selection operator that repairs destination support without retraining the host predictor or increasing the maintained set size. At a repair step, DSR evaluates a temporary destination-stratified candidate bank from the observed prefix, converts candidate evidence into integer target counts, protects representatives of active modes, and reallocates redundant surplus hypotheses to deficient modes. The maintained and returned sets retain exactly N hypotheses, and DSR replaces at most $\lceil \rho N \rceil$ entries. Protected representatives preserve current categorical support; lineage-aware particle filters also preserve surviving resampling ancestors. Each replacement reduces the allocation mismatch to the evidence-driven target by one. On the complete 3,719-trajectory Edinburgh protocol over three seeds, DSR reduces MIF weighted ADE and FDE by 13.36% and 13.30% at $N = 64$. Paired integrations with CLiFF, PPT, causal GDTS, Social Informer, and PECNet improve both metrics in every evaluated pair. These results show that finite-set support allocation is a useful prediction-side control point when a fixed hypothesis set serves as the interface to downstream systems.

I. INTRODUCTION

Mobile robots operating around pedestrians must anticipate more than one plausible human future. In practice, a prediction module often exposes a finite hypothesis set to downstream decision making, so the number of represented futures is itself a limited computational resource. If a plausible destination disappears from that set, a downstream planner cannot reason over that alternative even when it remains compatible with the observations. This makes the allocation of a fixed prediction budget relevant in its own right, in addition to the quality of the underlying trajectory generator.

Modern pedestrian predictors represent multimodality with recurrent interaction models, graph architectures, endpoint-conditioned generation, latent variables, and diffusion models [1]–[8]. Recent methods further improve destination

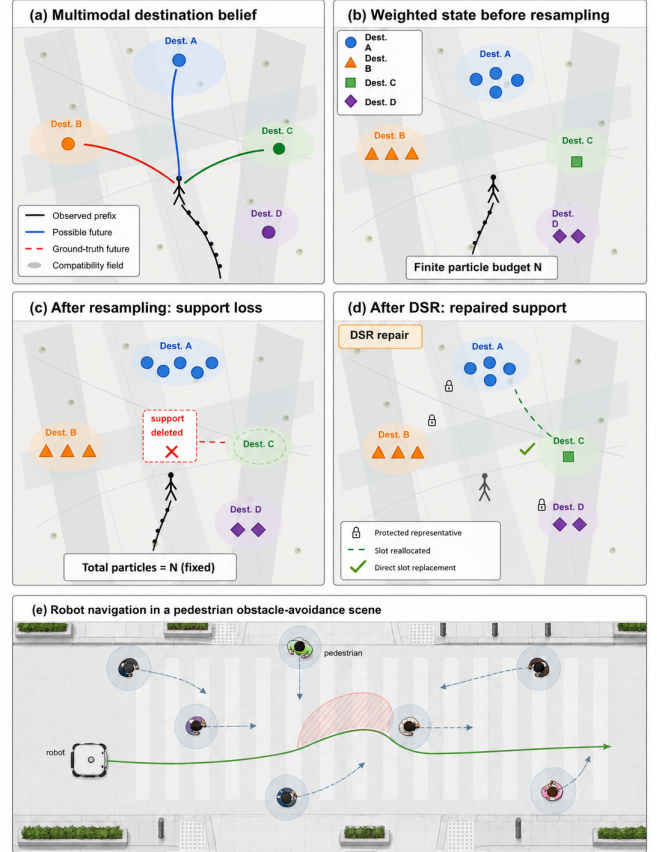


Fig. 1. Overview of Destination Support Restoration (DSR). (a) An observed prefix remains compatible with multiple destination modes. (b) A fixed-size maintained set concentrates on dominant modes. (c) Resampling may eliminate a still-plausible destination from the finite representation. (d) DSR protects surviving support and reallocates surplus slots toward evidence-supported deficits while returning the same number of hypotheses. (e) The figure illustrates the lineage-aware case; set-based hosts use protected mode representatives instead.

learning, goal-guided diffusion, and probabilistic Transformer prediction [9]–[11]. These advances improve how candidate futures are generated, but they do not by themselves decide how a finite maintained set should allocate its slots across plausible modes.

Fig. 1 shows the failure that motivates this work. A fixed-size set can concentrate on a dominant destination while assigning too few hypotheses, or none, to another plausible destination. The problem is especially visible in recursive predictors: repeated propagation, weighting, and resampling can replicate high-weight particles and remove low-mass categorical destinations from the finite representation [12],

¹Fengrui Liu is with the School of Computer Science and Technology, East China Normal University, Shanghai, China.

²Jiajun Peng is with the School of Data Science, University of Science and Technology of China, Hefei, China.

³Duo Peng is with the School of Computer Science and Technology, Tongji University, Shanghai, China.

⁴Feng Liu is with the School of Psychology, Shanghai Jiao Tong University, Shanghai, China.

*Corresponding author: Feng Liu.

[13]. Once a destination label disappears, continuous perturbations cannot recreate that categorical support. Goal-directed pedestrian filters and MIF-WLSTM therefore expose a gap between the multimodal belief and the support actually represented by the hypotheses available to the trajectory decoder [14]–[16].

Complete mode extinction is only the limiting case. A plausible destination may still retain one hypothesis while receiving much less of the fixed budget than the current evidence warrants. We refer to both cases as a **support-allocation mismatch**: the finite set no longer reflects how its slots should be distributed across plausible alternatives. The goal is not to enlarge the output set, but to use the same N slots more effectively by restoring missing support and reallocating redundant support before extinction occurs.

Existing particle-filter remedies address related forms of sample impoverishment. Resample–move methods rejuvenate particles through Markov transitions, KLD sampling changes the particle count, and alternative resampling rules redistribute particle mass [17]–[21]. Random mutation or restart may reintroduce a missing mode, but it does not assign an evidence-driven slot target to each semantic alternative. Diverse trajectory samplers improve one-shot coverage of a frozen predictor [22], [23]; our concern is different: how an already finite maintained set should allocate its existing slots online.

We propose **Destination Support Restoration (DSR)**, a causal operator for this allocation problem. At a repair step, DSR builds a temporary mode-stratified candidate bank from the observed prefix, converts candidate evidence into integer target counts for the N maintained slots, protects representatives of active support, and releases only unprotected hypotheses from target-surplus modes. Donors and recipient candidates are ranked by compatibility energy, and at most $\lceil \rho N \rceil$ slots are reassigned. The host predictor, maintained set size, and returned set size remain unchanged.

DSR separates support repair from the host generator. In a lineage-aware filter such as MIF-WLSTM [16], it protects the first offspring of every distinct resampling ancestor. Predictors without ancestry protect one representative of every active mode. The same allocation rule acts directly on MIF particles, through GoalBridge for CLiFF-LHMP [24], and through mode adapters for PPT [9], GDTS [10], Social Informer [11], and PECNet [5]. This lets the prediction-side repair sit between different generators and the finite hypothesis sets exposed downstream.

On the complete Edinburgh online protocol with 3,719 test trajectories, DSR reduces MIF weighted ADE and FDE by **13.36% and 13.30%** with the same 64 maintained and returned hypotheses. Paired host integrations also reduce both metrics for CLiFF, PPT, causal GDTS, Social Informer, and PECNet. The gains therefore do not depend on a single recursive filter or trajectory generator.

This paper makes three contributions:

- We identify support allocation under a fixed hypothesis budget as a distinct failure mode of multimodal prediction, covering both complete mode loss and under-

allocation before extinction.

- We introduce DSR, a bounded causal repair rule that turns current candidate evidence into target mode counts, preserves active support, and moves surplus slots toward deficient modes. We establish support preservation, bounded intervention, and exact contraction toward the target allocation.
- We evaluate the same repair principle across a destination-particle filter, a flow-map bridge, and learned finite-set predictors, with cross-scene, component, support-event, sensitivity, and runtime analyses to separate where the gains come from and what they cost.

II. RELATED WORK

A. Multimodal and Goal-Conditioned Trajectory Prediction

Pedestrian forecasting has progressed from recurrent social-interaction models to increasingly expressive multimodal generators. Social LSTM models interactions through recurrent social pooling [1], while Social GAN introduces adversarial generation to produce multiple socially plausible futures [2]. Trajectron++ combines dynamic interaction graphs, latent variables, and agent dynamics [3], and SGCN learns sparse spatial–temporal interaction graphs [4]. More recent diffusion models, including MID and LED, model stochastic futures through iterative denoising and provide expressive multimodal trajectory distributions [7], [8].

A complementary line of work structures long-horizon uncertainty around endpoints or goals. PECNet first predicts a distant endpoint and conditions the intermediate trajectory on that endpoint [5], while Y-Net decomposes long-term forecasting into goals, waypoints, and paths [6]. PPT similarly incorporates destination prediction into a progressive learning framework, using short-term and destination pretext tasks before full trajectory prediction [9]. GDTS explicitly uses estimated goals to guide diffusion and introduces tree sampling for efficient multimodal generation [10]. Social Informer instead combines an Informer-based predictor with interaction modeling and an adaptive variance mechanism to represent stochastic pedestrian motion [11].

These methods improve how a model learns or samples a conditional trajectory distribution. DSR acts after the host exposes modes and conditional candidates. It reallocates the maintained and returned hypotheses across those modes. Goal-conditioned and multimodal predictors therefore serve as DSR hosts, while DSR leaves their generators unchanged.

B. Recursive Intention and Long-Horizon Prediction

Recursive probabilistic predictors maintain temporal continuity by updating latent motion intent as observations arrive. Goal-Directed Pedestrian Prediction represents destinations as latent variables and recursively estimates their probabilities [14], while multi-hypothesis filtering maintains alternative pedestrian intentions during sequential inference [15]. MIF-WLSTM combines a mutable intention filter over pre-defined destination regions with a Warp LSTM trajectory decoder [16]. Its mutation mechanism can introduce random replacements, but it does not construct an evidence-derived

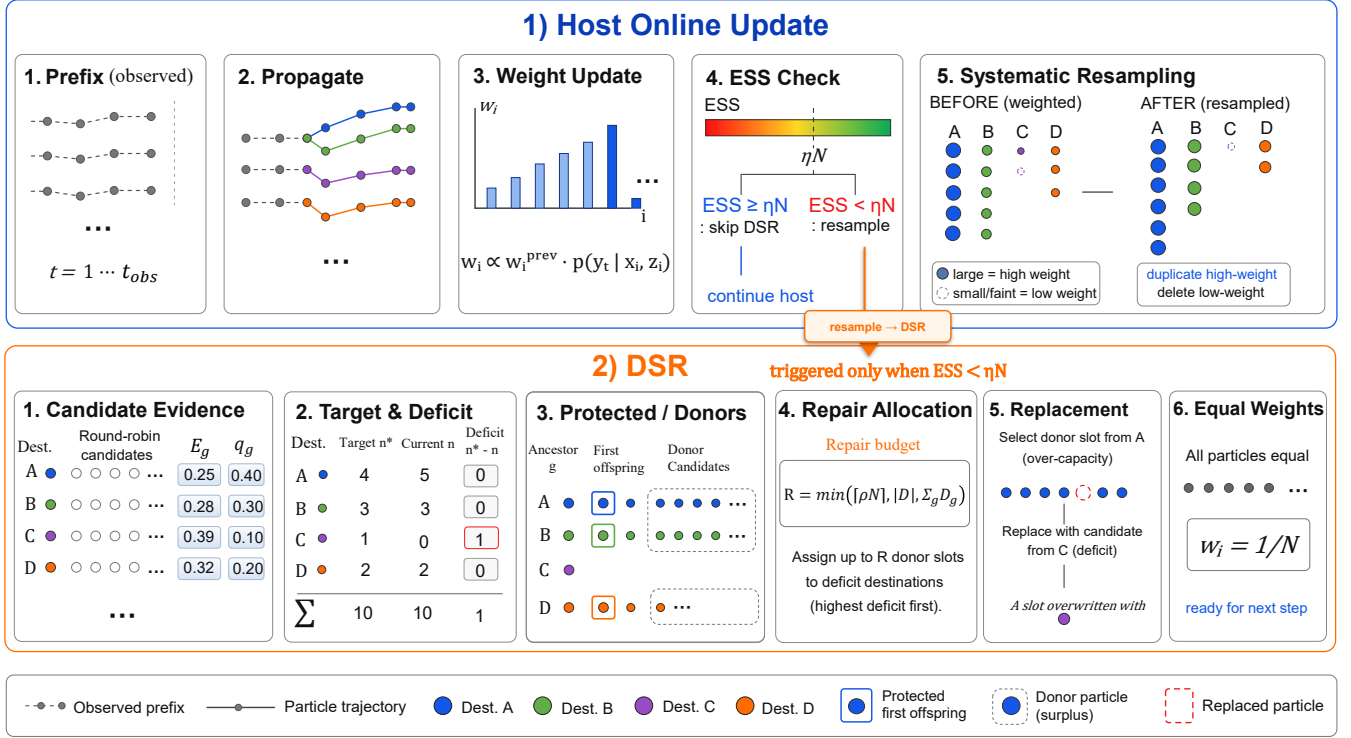


Fig. 2. Host update and DSR pipeline. The host first propagates and weights its maintained hypotheses and then checks the effective sample size. When $ESS < \eta N$, DSR maps evidence from a temporary destination-stratified candidate bank to target counts, protects active-mode representatives, replaces only unprotected surplus hypotheses, and returns N equally weighted hypotheses to the host.

target count for each destination, restrict donation to target-surplus hypotheses, or explicitly protect every surviving resampling lineage.

CLIFF-LHMP samples long-horizon motion from learned flow-field dynamics [24]. CLIFF exposes no particle ancestry, so our integration keeps its generated paths outside the DSR particle state. DSR instead maintains a finite goal-support state, and GoalBridge uses the repaired distribution to reweight the original CLIFF proposals. Lineage-aware filters provide ancestry-based protection; other predictors expose host-defined modes and protected representatives.

The focus of DSR is therefore narrower than constructing a complete recursive predictor. Propagation, observation processing, resampling or hypothesis generation, and trajectory decoding remain host operations. DSR intervenes only in the finite allocation between those stages.

C. Particle Rejuvenation, Resampling, and Diverse Hypothesis Selection

The finite-sample behavior motivating DSR connects to sample impoverishment in Sequential Monte Carlo [12], [13]. Auxiliary particle filters use look-ahead information to improve proposal selection [25], resample-move methods apply Markov transitions after resampling [17], and KLD sampling adapts particle count to approximation complexity [18]. Mode-preserving and deterministic resampling retain distributional structure or reduce impoverishment [19], [20]. Entropy-regularized optimal-transport resampling differentially redistributes particle mass [21].

DSR differs in both intervention point and objective. It leaves the host propagation, likelihood update, ESS criterion, and resampling rule unchanged. After the host has produced a finite set, DSR treats the number of slots assigned to each semantic mode as an explicit resource. It first protects the support that must survive, releases only target-surplus hypotheses outside that protected set, and reallocates a bounded number of those slots toward evidence-supported deficits. The resulting operation is not presented as an invariant Markov transition or an unbiased posterior correction; it is a controlled repair of the finite representation.

Output-diversification methods address another related problem. Likelihood-Based Diverse Sampling searches for separated high-likelihood trajectories from a pretrained distribution [22], while NPSN learns purposive samples for a frozen stochastic predictor [23]. Such methods improve coverage when constructing a new output set from a fixed observation window. DSR operates on an instantiated finite support state and reasons about categorical counts, protected representatives, and surplus-to-deficit reallocation. Diverse sampling controls where a generator places hypotheses; DSR controls their semantic allocation in the maintained set.

III. METHOD

Fig. 2 summarizes the complete intervention between the host resampling step and the next online update.

A. Finite Sets Expose Allocation Error

At each prediction step, the host maintains N hypotheses

$$\mathcal{S} = \{(x_i, z_i)\}_{i=1}^N, \quad (1)$$

where x_i is a continuous state and $z_i \in \{1, \dots, G\}$ is a destination or host-defined mode. Let

$$n_g = \sum_{i=1}^N \mathbf{1}[z_i = g] \quad (2)$$

be the number of hypotheses in mode g .

DSR redistributes the N maintained slots across plausible modes. It requires a mode-conditioned candidate sampler $Q(x | g)$, a causal compatibility energy $E(x, g; y_{1:t})$, and a protected set. For particle-filter hosts, DSR protects the first offspring of each surviving resampling ancestor. Hosts without ancestry protect one representative of every active mode. This second interface guarantees categorical support preservation; it makes no ancestry claim.

B. Candidate Evidence Defines Target Counts

At a repair step, DSR draws a temporary candidate bank stratified across modes. The bank exists only for the current allocation decision; DSR discards every candidate that does not enter the maintained set. For candidates $\mathcal{C}_g$ assigned to mode g , we compute

$$s_g = \text{LSE}_{c \in \mathcal{C}_g} [-\beta E(c, g; y_{1:t})], \quad \mathbf{q} = \text{softmax}(\mathbf{s}). \quad (3)$$

Here q_g measures the current evidence for allocating finite support to mode g ; it is not assumed to be a calibrated destination posterior.

We convert the real-valued allocation $N\mathbf{q}$ to integer target counts n_g^* with largest-remainder rounding, such that

$$\sum_{g=1}^G n_g^* = N, \quad |n_g^* - Nq_g| < 1. \quad (4)$$

Only plausible underrepresented modes are eligible for repair:

$$d_g = (n_g^* - n_g)_+ \mathbf{1}[q_g \geq \tau], \quad (5)$$

while a mode can donate at most

$$c_g = (n_g - n_g^*)_+ \quad (6)$$

hypotheses.

C. Surplus Donation Preserves Support

DSR releases slots only from **unprotected surplus hypotheses**. It ranks eligible donors by decreasing energy and removes the least prefix-compatible hypotheses first. We bound the number of replacements by

$$R = \min \left(\lceil \rho N \rceil, |D|, \sum_{g=1}^G d_g \right), \quad (7)$$

where D is the eligible donor set and ρ is the maximum repair ratio.

The greedy allocator assigns the R slots to deficient modes. With r_g replacements already assigned to mode g , it assigns the next slot to

$$g^* = \arg \max_{g: r_g < d_g} q_g d_g (d_g - r_g). \quad (8)$$

This favors modes that are both plausible and below their target allocation.

DSR ranks candidates within each recipient mode by increasing energy and directly replaces the selected donors. The pair (q_g, τ) determines mode plausibility, while energy orders donor removal and candidate insertion. For particle-filter hosts, DSR operates on the finite set produced after the host likelihood update and resampling.

D. Repair Contracts the Allocation Error

Two finite-set properties follow from protected donor selection and surplus-to-deficit replacement.

Proposition 1 (Support safety and bounded intervention). DSR modifies at most $\lceil \rho N \rceil$ hypotheses. If the protected set contains at least one representative of every active mode, then

$$\text{supp}_z(\mathcal{S}) \subseteq \text{supp}_z(\mathcal{S}'). \quad (9)$$

For lineage-aware particle filters, protecting the first offspring of every distinct resampling ancestor additionally preserves at least one unchanged offspring of each surviving ancestor.

Proposition 2 (Allocation contraction). Define

$$\Phi(\mathbf{n}) = \frac{1}{2} \|\mathbf{n} - \mathbf{n}^*\|_1. \quad (10)$$

Because every replacement moves one slot from a target-surplus mode to a target-deficient mode,

$$\Phi(\mathbf{n}') = \Phi(\mathbf{n}) - R. \quad (11)$$

Thus every DSR intervention moves the finite categorical allocation exactly R steps toward its evidence-driven target.

MIF-WLSTM provides the lineage-aware instance through its destination particles and resampling ancestors. CLiFF uses a repaired goal-support state to guide its original proposals. PPT, GDTs, Social Informer, and PECNet use the support-safe interface: their adapters assign host hypotheses to modes and protect one representative of each active mode. Each host defines its candidate generator and compatibility energy; all hosts share the DSR allocation operator.

IV. EXPERIMENT

A. Experimental Setup

1) *Dataset and Online Protocol:* We evaluate DSR on the Edinburgh Informatics Forum dataset [29] with the online prediction protocol released with MIF-WLSTM [16]. We represent trajectories at 10 Hz and use 34 square $1.5 \text{ m} \times 1.5 \text{ m}$ boundary regions as destination categories. The test split contains 3,719 trajectories, and each run produces 86,404 online updates. Both prediction and correction windows contain 20 positions. Prediction starts after a 10-position warm-up and advances every two positions.

TABLE I

RESULTS ON EDINBURGH. LOWER IS BETTER. WE INCLUDE HISTORICAL AOE/FOE VALUES MARKED WITH $\dagger$ AS CONTEXT BECAUSE THEY FOLLOW A DIFFERENT PROTOCOL. NLL COMPARISONS ARE PAIRED WITHIN HOST; FOR CLIFF, WE MEASURE THE DSR REDUCTION RELATIVE TO GOALBRIDGE.

Predictor	Configuration	ADE/AOE $\downarrow$	FDE/FOE $\downarrow$	NLL $\downarrow$	DSR Reduction
Social Force [16], [26] $\dagger$	-	3.1240	3.9090	-	-
LSTM [16], [27] $\dagger$	-	2.1320	3.0050	-	-
Social LSTM [1], [16] $\dagger$	-	1.5240	2.5100	-	-
Attention LSTM [16], [28] $\dagger$	-	0.9860	1.3110	-	-
Social GAN [2], [16] $\dagger$	340 outputs	1.0420	2.0880	-	-
MIF [16]	$N = 64$	0.6594	1.2420	5.1041	-
MIF [16]	$N = 340$	0.6051	1.1404	3.9635	-
MIF + DSR	$N = 64$	0.5713	1.0768	3.7024	13.36% / 13.30%
CLiFF [24]	$K = 64$	0.5542	1.1194	3.2814	-
CLiFF + GoalBridge	$K = 64$	0.5319	1.0638	3.2021	-
CLiFF + GoalBridge + DSR	$K = 64$	0.5264	1.0493	3.1481	1.03% / 1.36%
PPT [9]	$H = 20$	1.0931	2.4961	1.5351	-
PPT + DSR	$H = 20$	1.0024	2.2842	1.4359	8.30% / 8.49%
Causal GDTs [10]	$H = 20$	0.4925	0.9615	1.0163	-
Causal GDTs + DSR	$H = 20$	0.4563	0.8947	0.9853	7.36% / 6.95%
Social Informer [11]	$H = 20$	0.4951	0.9683	1.0100	-
Social Informer + DSR	$H = 20$	0.4513	0.8874	0.9765	8.85% / 8.36%
PECNet [5]	$H = 20$	0.4916	1.0554	1.8110	-
PECNet + DSR	$H = 20$	0.3709	0.7675	2.3484	24.55% / 27.27%

We report three-seed means for every method in the unified protocol. For each host, we compare the same model and prediction configuration with and without DSR. The evaluation emphasizes paired within-host changes because the host architectures use different probability parameterizations. All comparative conclusions reported below are supported by paired statistical analysis; complete confidence intervals and test outputs will be released together with the code.

MIF uses $N = 64$ maintained destination particles in the main comparison. CLiFF returns $K = 64$ proposals and connects to the repaired destination state through GoalBridge. PPT [9], causal GDTs [10], Social Informer [11], and PECNet [5] use the support-safe interface. Their adapters associate host hypotheses with endpoint or destination modes and protect one representative of each active mode.

The main MIF configuration uses $C = 68$, $\beta = 10$, $\tau = 0.01$, and $\rho = 0.10$. Candidate evidence determines the target allocation, energy ranks surplus donors and recipient candidates, and each repair replaces at most seven of the 64 maintained hypotheses. Host-specific adapters retain the same allocation rule.

2) *Metrics*: We report weighted average displacement error (wADE) and weighted final displacement error (wFDE),

$$\text{wADE} = \sum_{k=1}^K \omega_k \frac{1}{H} \sum_{h=1}^H \left\| \hat{\mathbf{y}}_h^{(k)} - \mathbf{y}_h \right\|_2, \quad (12)$$

$$\text{wFDE} = \sum_{k=1}^K \omega_k \left\| \hat{\mathbf{y}}_H^{(k)} - \mathbf{y}_H \right\|_2, \quad (13)$$

where we normalize the output weights ω_k before evaluation. We report negative log-likelihood (NLL) as a secondary probabilistic score and compare it only within each host because the predictors use different probability parameterizations. NLL is evaluated under each host's own post-integration output distribution. Because DSR targets finite-set support allocation rather than the host's probabilistic scoring objective, lower displacement error need not imply a lower NLL.

B. DSR Improves Paired Hosts

Table I compares DSR with its paired host baselines. We mark historical MIF-WLSTM results with $\dagger$ and include them only as context because they follow another protocol. We report relative DSR reductions only for matched host pairs.

1) *DSR Improves the Lineage-Aware Host*: On MIF, DSR reduces wADE from 0.6594 to 0.5713 and wFDE from 1.2420 to 1.0768 with the same $N = 64$ maintained and returned particles. These changes equal relative reductions

of 13.36% and 13.30%. The repaired $N = 64$ model also outperforms the $N = 340$ MIF reference, whose wADE and wFDE are 0.6051 and 1.1404. Complete DSR therefore outperforms a larger maintained set under the proposal and computation costs reported below.

2) *Complete DSR Transfers Across Hosts*: The improvement transfers beyond lineage-aware MIF. PPT reduces wADE and wFDE by 8.30% and 8.49%. Causal GDS improves by 7.36%/6.95%, while Social Informer improves by 8.85%/8.36%. These models expose no resampling ancestry, so their adapters protect mode representatives. The paired gains show that complete DSR transfers across finite-set predictors with different architectures.

PECNet shows the largest displacement improvement, reducing wADE from 0.4916 to 0.3709 and wFDE from 1.0554 to 0.7675, corresponding to reductions of 24.55% and 27.27%. The complete NLL column decreases for MIF, CLiFF+GoalBridge, PPT, causal GDS, and Social Informer; PECNet is the sole paired exception, increasing from 1.8110 to 2.3484. This divergence is consistent with DSR’s objective: reallocating a finite support set can improve geometric trajectory coverage without necessarily improving the host-specific probabilistic score.

For CLiFF, the strict comparison is between GoalBridge and GoalBridge+DSR. DSR further reduces wADE/wFDE by 1.03%/1.36%. We do not attribute the larger difference between the original CLiFF output and GoalBridge+DSR entirely to DSR because GoalBridge itself modifies the proposal weights.

C. Cross-Scene Generalization

Table II evaluates MIF, CLiFF, and PPT on the five ETH/UCY scenes. Each entry reports a three-seed mean from a paired frozen-host comparison. MIF and CLiFF follow the destination-filter online protocol, while PPT uses its official frozen scene checkpoints with shared raw-track 8-observation/12-future windows. We compare each host only with its paired DSR integration because these protocols induce different absolute error scales.

DSR lowers wFDE in all 15 host-scene pairs and lowers mean wADE in 14. The sole exception is PPT on ZARA1, where wADE changes from 2.2856 to 2.2858 while wFDE falls by 6.02%. PPT yields a 5.16% macro-average wFDE reduction across the five scenes. These results extend the Edinburgh finding across new scenes and two host interfaces.

D. Ablation of DSR Components

We ablate four design choices on MIF with $N = 64$ while keeping the remaining evaluation configuration fixed. Setting $\rho = 0$ disables repair entirely. The second variant removes lineage protection together with the surplus-only donor constraint. The final two variants replace the energy-based donor ranking or candidate ranking with random selection, respectively.

Disabling repair causes by far the largest deterioration, increasing wADE/wFDE by 21.46%/21.35% relative to Full DSR, which confirms that support repair is the

TABLE II
PAIRED ETH/UCY RESULTS OVER THREE SEEDS. EACH CELL REPORTS
HOST→DSR; LOWER IS BETTER. UNIV DENOTES `stu03` FOR MIF
AND CLiFF.

Host	Scene	wADE ↓		wFDE ↓	
MIF	ETH	1.2456 →	1.2255	2.6604 →	2.5983
	HOTEL	0.4133 →	0.4066	0.9036 →	0.8825
	UNIV	0.7292 →	0.7277	1.5781 →	1.5699
	ZARA1	0.5258 →	0.5185	1.1513 →	1.1306
	ZARA2	0.4222 →	0.4171	0.9280 →	0.9134
CLiFF	ETH	1.1324 →	1.1205	2.2120 →	2.1076
	HOTEL	0.4761 →	0.4565	0.9777 →	0.9683
	UNIV	0.7668 →	0.7546	1.6127 →	1.6062
	ZARA1	0.5585 →	0.5536	1.1591 →	1.1492
	ZARA2	0.4396 →	0.4287	0.9193 →	0.9130
PPT	ETH	4.2873 →	4.2848	4.3242 →	4.2958
	HOTEL	1.8756 →	1.8532	1.8856 →	1.7762
	UNIV	1.6133 →	1.6001	2.1627 →	2.0534
	ZARA1	2.2856 →	2.2858	1.8381 →	1.7274
	ZARA2	1.5583 →	1.5396	1.7013 →	1.5604

TABLE III
COMPONENT ABLATION ON MIF WITH $N = 64$. LOWER IS BETTER.

Variant	wADE ↓	wFDE ↓
Full DSR	0.5713	1.0768
Zero repair ($\rho = 0$)	0.69391	1.30667
w/o lineage + surplus protection	0.58088	1.09413
Random donor	0.59654	1.09956
Random candidate	0.58372	1.09875

main source of the gain. Removing lineage protection together with surplus-only donation also degrades performance to 0.58088/1.09413, indicating that the two protection mechanisms are effective jointly. Replacing energy-based donor ranking with random donor selection increases wADE/wFDE to 0.59654/1.09956, corresponding to degradations of 4.42%/2.11%. Random candidate selection yields 0.58372/1.09875, or 2.17%/2.04% worse than Full DSR. These results show that energy-based ranking contributes beyond mode-level repair, with donor ranking having the larger effect on average trajectory accuracy.

E. Displacement Gains Persist After Support Repair

Three support events test whether complete DSR behaves as the intended support-repair mechanism. Immediate revival records a mode that regains support at the repair step. Historical recovery records a mode that had zero support earlier and becomes active again. Pre-extinction increase records additional support assigned to a mode before it disappears. Random Reset is matched to DSR in intervention trigger and replacement count R , but selects donor slots and recipient candidates uniformly at random from the corresponding eligible pools rather than using target-deficit allocation or

TABLE IV

SUPPORT-REPAIR RATES ON MIF WITH $N = 64$. GAIN IS THE CLUSTER-PAIRED DIFFERENCE BETWEEN METHOD-SPECIFIC RATES IN PERCENTAGE POINTS (PP).

Mechanism	Random Reset	DSR	Gain (pp)
Immediate revival	2.94%	24.11%	21.17
Historical recovery	3.15%	22.21%	19.07
Pre-extinction increase	1.98%	22.72%	20.74

energy-based ranking.

Table IV uses 11,157 seed-trajectory clusters and 2,000 bootstrap repetitions. We use the true destination only to label posthoc support events; DSR receives no future labels. Each method induces its own eligible-event set, and the paired bootstrap matches seed-trajectory clusters instead of individual events. DSR raises every support-repair rate by at least 19.07 pp over Random Reset. Active-mode stratification localizes the displacement gains: DSR reduces wADE by 13.69% with 1–2 active modes, 16.47% with 3–4, and 4.96% with at least five. Complete DSR therefore improves every bucket and yields its largest reduction when several alternatives compete for a small maintained set.

The intervention remains sparse. DSR invokes on 22.33% of updates and performs a nonzero repair on 21.63% of all updates. An active repair replaces a mean of 6.737 hypotheses and a median of seven, or 10.53% and 10.94% of the maintained set. Later predictions retain lower displacement error: after immediate revival, the mean DSR-minus-host wADE remains -0.0784 at +1 update, -0.1056 at +5, and -0.1257 at +10. wFDE follows the same pattern at -0.1476 , -0.2011 , and -0.2428 . Paired cluster-bootstrap intervals support the same conclusions; the complete intervals and test outputs will be released with the code.

F. Sensitivity to Repair Capacity and Support Threshold

Fig. 3 varies the repair ratio ρ , candidate budget C , and support threshold τ on MIF. The figure reports the relative reduction against the default setting; positive values favor the varied setting. Setting $\rho = 0$ removes repair and increases wADE, wFDE, and NLL by 21.5%, 21.4%, and 64.2% relative to the default. Increasing ρ from 0.10 to 0.20 and increasing C from 68 to 272 further reduce all three metrics, with higher runtime and more candidate scoring. The default values therefore balance intervention size and cost instead of maximizing offline accuracy. Results remain nearly unchanged for $\tau \in [0, 0.01]$, while $\tau = 0.02$ raises wADE and wFDE by about 0.2%.

G. Accuracy–Runtime Trade-off

Fig. 4 summarizes the accuracy–runtime trade-off of DSR on three representative Edinburgh hosts. For MIF, average inference runtime increases from 30.823 ms to 35.523 ms per update (15.25%), while wADE and wFDE decrease by 13.36% and 13.30%. For CLiFF+GoalBridge, runtime increases from 35.807 ms to 38.758 ms (8.24%), with corresponding wADE/wFDE reductions of 1.03%/1.36%. PECNet

has a much lighter baseline: runtime rises from 0.90 ms to 6.13 ms per step, an absolute increase of 5.23 ms, while wADE/wFDE decrease by 24.55%/27.27%.

V. DISCUSSION AND CONCLUSION

The experiments support the central premise of DSR: under a fixed prediction budget, explicitly repairing support allocation improves the finite trajectory set without changing the host generator. The component ablation identifies repair as the dominant source of gain, the support-event analysis links the improvement to the intended restoration mechanism, and the cross-scene and runtime results show how the effect transfers and what additional computation it requires. More broadly, DSR shows that finite-set support allocation can serve as an explicit, lightweight control layer between prediction and downstream decision making.

DSR remains a prediction-side intervention rather than a complete planning solution, and its effectiveness depends on meaningful mode definitions and compatible candidate generation. Our evaluation is limited to pedestrian trajectory prediction with fixed repair hyperparameters; closed-loop robot experiments are left for future work. Within that scope, DSR provides a practical way to make the finite hypothesis set passed to downstream decision making better represent plausible human futures without increasing its maintained size.

REFERENCES

- [1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, “Social LSTM: Human trajectory prediction in crowded spaces,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 961–971.
- [2] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, “Social GAN: Socially acceptable trajectories with generative adversarial networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2255–2264.
- [3] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, “Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data,” in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 683–700.
- [4] L. Shi, L. Wang, C. Long, S. Zhou, M. Zhou, Z. Niu, and G. Hua, “SGCN: Sparse graph convolution network for pedestrian trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 8994–9003.
- [5] K. Mangalam, H. Girase, S. Agarwal, K.-H. Lee, E. Adeli, J. Malik, and A. Gaidon, “It is not the journey but the destination: Endpoint conditioned trajectory prediction,” in *Proceedings of the European Conference on Computer Vision*, 2020, pp. 759–776.
- [6] K. Mangalam, Y. An, H. Girase, and J. Malik, “From goals, waypoints & paths to long term human trajectory forecasting,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 233–15 242.
- [7] T. Gu, G. Chen, J. Li, C. Lin, Y. Rao, J. Zhou, and J. Lu, “Stochastic trajectory prediction via motion indeterminacy diffusion,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 113–17 122.
- [8] W. Mao, C. Xu, Q. Zhu, S. Chen, and Y. Wang, “Leapfrog diffusion model for stochastic trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5517–5526.
- [9] X. Lin, T. Liang, J. Lai, and J.-F. Hu, “Progressive pretext task learning for human trajectory prediction,” in *Proceedings of the European Conference on Computer Vision*, 2024.
- [10] G. Sun, S. Wang, L. Zhu, M. Liu, and J. Ma, “GDTS: Goal-guided diffusion model with tree sampling for multi-modal pedestrian trajectory prediction,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2025.

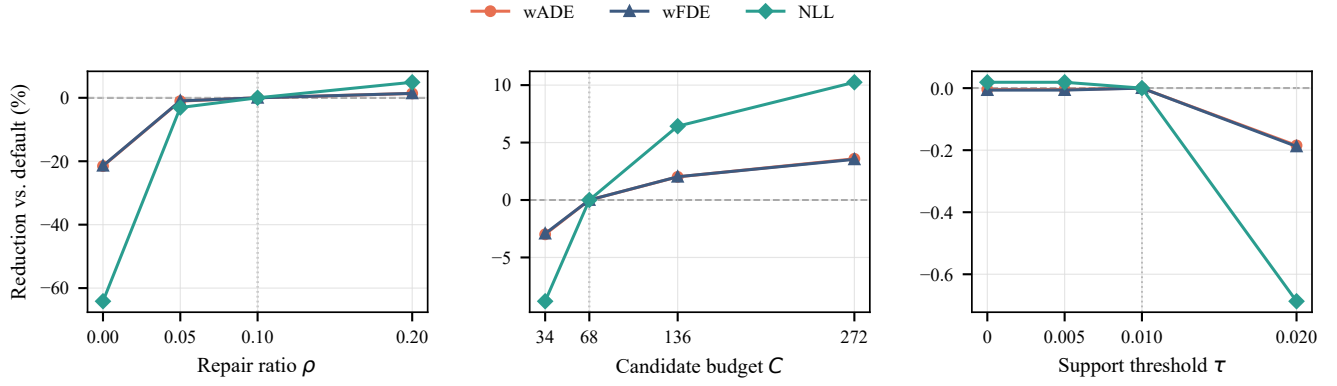


Fig. 3. Sensitivity on MIF Edinburgh. Each panel changes one parameter while fixing the others at $\rho = 0.10$, $C = 68$, and $\tau = 0.01$. Values show reduction relative to the default, so positive values indicate lower error. Dotted vertical lines mark the default settings.

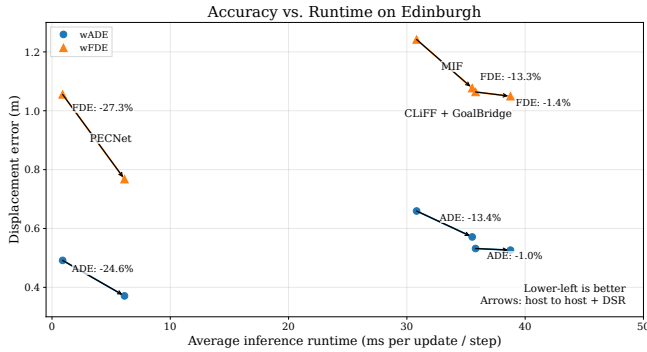


Fig. 4. Accuracy–runtime trade-off on Edinburgh for MIF, CLiFF+GoalBridge, and PECNet. Each arrow connects a host configuration to its DSR-enhanced counterpart; lower-left is better. DSR reduces both wADE and wFDE for all three hosts while increasing average inference runtime by a model-dependent amount.

- [11] Z. Jiang, R. Yang, Y. Ma, C. Qin, X. Chen, and Z. Wang, “Social informer: Pedestrian trajectory prediction by informer with adaptive trajectory probability region optimization,” *IEEE Transactions on Cybernetics*, vol. 56, no. 1, pp. 15–28, 2026.
- [12] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, “Novel approach to nonlinear/non-gaussian bayesian state estimation,” *IEE Proceedings F—Radar and Signal Processing*, vol. 140, no. 2, pp. 107–113, 1993.
- [13] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York, NY, USA: Springer, 2001.
- [14] E. Rehder and H. Kloeden, “Goal-directed pedestrian prediction,” in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 139–147.
- [15] F. Particke, M. Hiller, C. Feist, and J. Thielecke, “Improvements in pedestrian movement prediction by considering multiple intentions in a multi-hypotheses filter,” in *Proceedings of the IEEE/ION Position, Location and Navigation Symposium*, 2018, pp. 209–215.
- [16] Z. Huang, A. Hasan, K. Shin, R. Li, and K. Driggs-Campbell, “Long-term pedestrian trajectory prediction using mutable intention filter and warp LSTM,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 542–549, 2021.
- [17] W. R. Gilks and C. Berzuini, “Following a moving target—monte carlo inference for dynamic bayesian models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 1, pp. 127–146, 2001.
- [18] D. Fox, “KLD-sampling: Adaptive particle filters,” in *Advances in Neural Information Processing Systems*, vol. 14, 2001, pp. 713–720.
- [19] M. R. Morelande and A. M. Zhang, “A mode preserving particle filter,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 3984–3987.
- [20] T. Li, M. Bolić, and P. M. Djurić, “Deterministic resampling: Unbiased

- sampling to avoid sample impoverishment in particle filters,” *Signal Processing*, vol. 92, no. 7, pp. 1637–1645, 2012.
- [21] A. Corenflos, J. Thornton, G. Deligiannidis, and A. Doucet, “Differentiable particle filtering via entropy-regularized optimal transport,” in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 139, 2021, pp. 2100–2111.
- [22] Y. J. Ma, J. P. Inala, D. Jayaraman, and O. Bastani, “Likelihood-based diverse sampling for trajectory forecasting,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 279–13 288.
- [23] I. Bae, J.-H. Park, and H.-G. Jeon, “Non-probability sampling network for stochastic human trajectory prediction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6477–6487.
- [24] Y. Zhu, A. Rudenko, T. P. Kucner, L. Palmieri, K. O. Arras, A. J. Lilienthal, and M. Magnusson, “CLiFF-LHMP: Using spatial dynamics patterns for long-term human motion prediction,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2023, pp. 3795–3802.
- [25] M. K. Pitt and N. Shephard, “Filtering via simulation: Auxiliary particle filters,” *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.
- [26] D. Helbing and P. Molnár, “Social force model for pedestrian dynamics,” *Physical Review E*, vol. 51, no. 5, pp. 4282–4286, 1995.
- [27] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, “Soft+hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection,” *Neural Networks*, vol. 108, pp. 466–478, 2018.
- [29] B. Majecka, “Statistical models of pedestrian behaviour in the forum,” Master’s thesis, School of Informatics, University of Edinburgh, Edinburgh, U.K., 2009.