

Learnable Spatio-Temporal Map Embeddings for Deep Inertial Localization

Dennis Melamed, Karnik Ram, Vivek Roy, and Kris Kitani.

Abstract—Indoor localization systems often fuse inertial odometry with map information via hand-defined methods to reduce odometry drift, but such methods are sensitive to noise and struggle to generalize across odometry sources. To address the robustness problem in map utilization, we propose a data-driven prior on possible user locations in a map by combining learned spatial map embeddings and temporal odometry embeddings. Our prior learns to encode which map regions are feasible locations for a user more accurately than previous hand-defined methods. This prior leads to a 49% improvement in inertial-only localization accuracy when used in a particle filter. This result is significant, as it shows that our relative positioning method can match the performance of absolute positioning using bluetooth beacons. To show the generalizability of our method, we also show similar improvements using wheel encoder odometry. Our code will be made publicly available[†].

I. INTRODUCTION

Existing approaches to using occupancy map information to improve odometry-only localization are often hand-defined, leading to sensitivity to odometry errors and implicit assumptions which do not always generalize across odometry sources. Methods like SLAM [1], map-matching [2], [3], and heuristic map constraints on motion [4], [5] show that maps are a rich information source to improve localization, but new methods are needed to more robustly utilize map information. We seek to extract more robust map information via a learned model to improve localization in odometry-only scenarios, particularly for indoor localization.

Inertial odometry utilizes inertial measurement units (IMUs) like those in smartphones (Fig. 4b) to localize agents. The accuracy of inertial odometry has seen significant improvements from deep methods [6]–[8] but still suffers from drift due to integration of small errors in relative motion estimates. This is unavoidable without introducing absolute constraints on user location, like those provided by maps. Introducing map information allows for drift correction without requiring additional sensors to provide absolute constraints like cameras, LiDARs, or fingerprinting systems [9].

Existing indoor localization systems using vision, LiDAR, and fingerprinting suffer from downsides that odometry-only methods avoid. Image or LiDAR-based localization via Simultaneous Localization and Mapping (SLAM) can be highly accurate [1], [10], but can require well-lit and highly featured environments, be power-hungry, and be constraining for human activities since these sensors always have to be

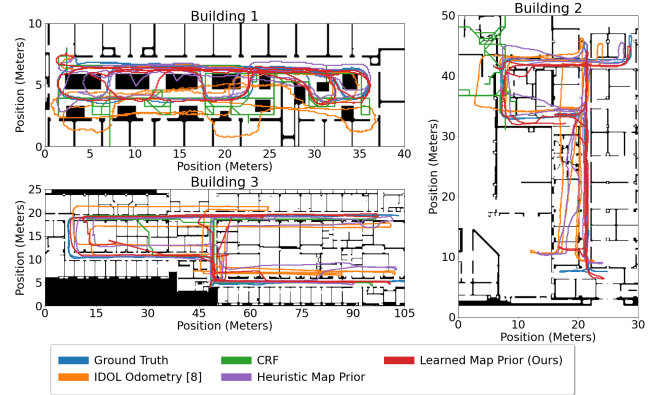


Fig. 1: Our proposed method utilizes learnable spatio-temporal map priors to reduce drift in inertial odometry. Plots show qualitative comparison of trajectories estimated by proposed method against baselines on different buildings (IDOL dataset [8]).

able to see the world. WiFi and Bluetooth fingerprinting systems [11], [12] require instrumentation of environments, which can be costly and require significant upkeep. The deep inertial odometry methods proposed in RoNIN [6], TLIO [7], and IDOL [8] avoid the above drawbacks by making use of IMUs. IMUs are lightweight, use only milliAmps of current, do not require line-of-sight, and are present in most modern smartphones and robots.

Map information has been utilized to reduce inertial odometry drift [3]–[5], but prior methods are generally hand-defined. These hand-defined methods make assumptions which generalize poorly, like discretizations of the map for human localization which are too coarse for slower-moving robots, or use heuristics on location feasibility. In our experimentation, these heuristics often lead to localization errors. To reduce odometry drift without hand-defined methods, our approach fuses spatial map information and temporal odometry information via an efficient learnable map prior. Our deep network uses a convolutional model to encode map information and a recurrent model to encode recent odometry measurements. These encodings are combined into a map prior to determine areas of the map where the recent odometry measurements would have been feasible with higher accuracy than prior methods. When used as the sensor model of a simple particle filter, our prior is able to utilize highly noisy inertial odometry to provide approximately 49% improvements in localization accuracy over the original odometry. This level of improvement is also seen when our system is applied to wheel encoder odometry on robots like in Fig. 4c without requiring any retraining.

All authors are with the Robotics Institute in Carnegie Mellon University, Pittsburgh, PA, USA. Correspondence: kkitani@cs.cmu.edu

[†]Project page: <https://rebrand.ly/learned-map-prior>

II. RELATED WORK

Indoor localization has been explored from several angles including pure odometry approaches, combining maps with odometry, and utilizing other sensors combined with odometry.

Pure Odometry Approaches: Inertial odometry has used both traditional and data-driven approaches. Traditional approaches generally use device heading estimation, step-detection, and heuristic stride length assumptions to perform Pedestrian Dead Reckoning (PDR) by adding up strides in the estimated user heading [13]. IONET [14] used a deep network to learn to consume IMU measurements and output user polar displacements. RoNIN [6] explored ResNets, temporal convolutional networks, and recurrent networks to determine the best performing architectures for the inertial localization task, and TLIO [7] added a Kalman filtering element to improve accuracy. IDOL [8] reduced inertial heading errors by using a learned deep orientation estimation system. Wheel encoders are also a common source of pure odometry, and recent applications of deep networks have provided better uncertainty estimation to reduce encoder odometry drift [15]. Despite improvements, most odometry-only approaches still suffer from drift over the long term.

Map & Odometry Fusion: Map information has been used for localization via several different techniques. Graph-based systems turn a map into a series of connected nodes which lie in the space the user might travel through. The user’s trajectory is then constructed out of physically feasible transitions between these nodes which best match the user’s odometry. Luo et al. [16] implement a hidden Markov model (HMM) to localize users’ unstable GPS signals on a graph of outdoor road segments. Deep attention networks have been shown to outperform HMMs in some scenarios for matching user travel to road segments [17]. Indoors, conditional random fields (CRFs) have seen high pedestrian localization accuracy by placing nodes in physically accessible space, spaced apart by human step lengths [3]. Transitions in this graph are predicted based on inertial odometry. Graph approaches are successful, but generally require hand-design of the underlying graph and transition model for a specific source of odometry or sensor measurement. For example, [3] implicitly assumes the desired localization precision is that of a human user’s step by their node spacing. A small wheeled robot may need more granular localization, meaning that the map graph has to be completely regenerated to transfer to the new platform.

Bayesian filtering systems improve localization accuracy by combining information sources like odometry with absolute constraints like maps or absolute position measurements. Extended Kalman filtering (EKF) is an efficient approach to estimate agent state from both odometry and external information [18], but suffers from only maintaining one estimate of state and assumes Gaussian uncertainty. Particle filters [19] are often used when state distributions are unknown, change over time, and could be multi-modal. Multiple estimates of location are propagated using odometry, then kept

or removed based on information from maps or other sensors. Particle filters are simple and adaptable so lend themselves well to localization on low compute devices. For particle filters, map constraints are usually invoked by heuristics which remove particles whose last odometry propagation is infeasible according to the map (e.g., penetrating a wall). Such a system has been proven useful in practical scenarios [4], [5], [13], but removing unlikely particles in this way is extremely sensitive to certain types of odometry noise (e.g., proximity to obstacles can remove too many particles). The particle filter presented by [2] computes overlap between map obstacles and a trajectory computed by integrating a few previous odometry measurements backwards from a particle location. Particles are given more weight if there are fewer overlaps. This type of multi-step heuristic is less sensitive to odometry noise than a single-step heuristic, but still suffers from the high levels of noise in inertial odometry. Our work suggests that a deep network may be able to capture more general shape characteristics of the odometry to better match map locations even when odometry error is high.

Localization with other sensors: Vision and LiDAR systems like V-LOAM [1] are highly accurate, but require line-of-sight to the environment and have significant power and computation needs. Bluetooth beacons or WiFi access points are also used to generate a pattern of radio frequency signals which can identify a location fairly precisely [9], [12]. Radio frequency systems are even more effective when combined with inertial odometry to smooth the noisy absolute location estimates [11]. However, such systems often require expensive infrastructure installation and maintenance, and building materials can block signal propagation. Utilizing an odometry-only approach resolves many of these problems, particularly if the sensors are low-power and non-line-of-sight like the IMUs and wheel encoders used here.

III. METHOD

We aim to develop a system which learns to consume high drift odometry measurements and occupancy map information to reduce drift in the estimated user location. Map information is integrated via the sensor model of a particle filter, but prior methods of doing so are sensitive to high levels of error in inertial odometry. Instead, we learn a data-driven prior from odometry and maps which we hypothesize will indicate likely user locations with higher accuracy than prior heuristic methods.

Model Architecture: Our learnable Map Prior network is shown in Fig. 2. The network aims to generate the likelihood score that a given series of odometry measurements ends at a given location \mathbf{x} in the map. The likelihood score can be used directly as the particle filter weight for a particle at \mathbf{x} . We set up this network with two branches to separate map and odometry processing, whose deep encodings are later combined. This setup is useful because maps are generally static, so at runtime a specific indoor map only needs to be processed by the network once to generate its encoding. The stored map encoding can later be combined with the constantly changing odometry encoding, reducing computational

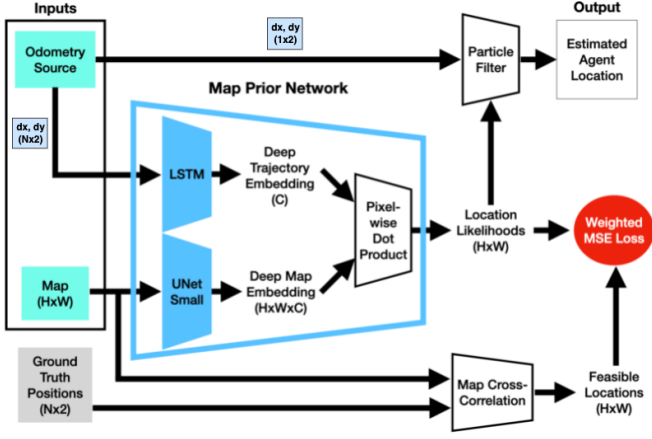


Fig. 2: System Diagram: Our map prior network takes as input (1) occupancy map and (2) window of odometry measurements, and outputs a map of likelihood scores. Scores represent map locations that are likely given input odometry measurements. Scores are used to weigh samples of a particle filter for tracking agent’s trajectory.

burden. The inputs to our method are a 2D indoor occupancy map M and a short history of the agent’s most recent odometry O from time $t-n$ to t . n is selected experimentally to be 5 seconds for human walking and 20 seconds for slower robot driving. Our map network $f(M, \Theta_f)$ is a small version of U-Net [20], [21] parameterized by weights Θ_f which has been modified to output a “Deep Map Tensor” of the same spatial dimensions as the input map, but with more channels. The number of channels $c = 32$ is chosen experimentally, balancing speed of computation with sufficient space to store relevant environmental details. The odometry network $g(O, \Theta_g)$ is a two-layer Long-Short-Term Memory (LSTM) network parameterized by weights Θ_g operating on the time dimension of the odometry history. The LSTM hidden state is of size c , and the last hidden state output is used as the “Deep Trajectory Vector”. A dot product is taken between the Deep Map Tensor and the Deep Trajectory Vector at each location \mathbf{x} , giving the likelihood score of \mathbf{x} given O and M :

$$S(\mathbf{x}|M, O) = f(M, \Theta_f)_{\mathbf{x}} \cdot g(O, \Theta_g) \quad (1)$$

Model Training: To train our model, we window ground truth walking trajectory data into windows of n timestamps, and subtract the initial position in each window to make this data look like odometry data. We additionally augment the training data and add a multiplicative velocity bias drawn from $\mathcal{N}(1, 0.5)$ pixel/second and a random noise drawn from $\mathcal{N}(0, 0.25)$ pixels to each input sample to better mimic noisy odometry data. These windows are used as odometry branch training input. The map branch is given small segments of the relevant map at train time to avoid overfitting to the most commonly feasible areas of the full map.

We create a target T (Fig. 3c) defining which areas of the map are feasible end-locations for a given ground-truth trajectory window. We train our network to output T , even when the odometry is given as input is noisy or drifting. A 2D kernel like in Fig. 3b is created from the ground truth trajectory segment and is cross-correlated with the map to

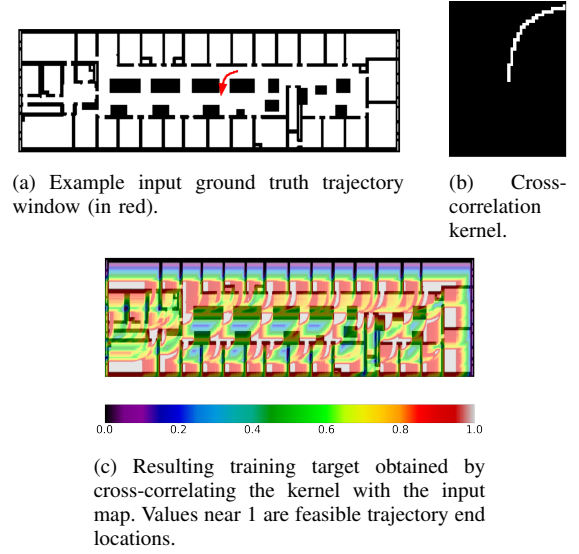


Fig. 3: Ground truth generation process for training our map prior network.

form \bar{T} whose values depend on the input trajectory’s overlap with obstacles. $T = 10^{-6} * e^{14*\bar{T}}$ is computed to heavily upweight areas where the trajectory window passes only through free space, *i.e.* is physically feasible. We train using Mean Squared Error (MSE) loss: $\mathcal{L} = \text{MSE}(T, S(\mathbf{x}|M, O))$. In training our method we noticed that smaller feasible regions of T were not being predicted well. We weight feasible regions by the inverse of their pixel area in our loss to improve small area prediction.

Particle Filter: We utilize a particle filter to combine the learned map prior with odometry measurements, as particle filters have relatively low computational burden but are able to handle complex distributions of estimated location. A particle filter maintains p estimates of the user’s state (location) as particles. At each timestep, each particle has its stored state updated according to the motion model:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \Delta \mathbf{x} + \eta_{\text{motion}} \quad (2)$$

where \mathbf{x}_t is the user’s state (x, y) at timestep t , and $\Delta \mathbf{x}$ is the odometry measurement between times t and $t + 1$. 2D Gaussian noise $\eta_{\text{motion}} \sim \mathcal{N}(\mu, \Sigma)$ is added to each particle’s motion to capture odometry uncertainty.

For our wheeled robot experiments, due to the significantly higher heading error present in wheel encoder odometry, we extend the state to also include robot heading θ :

$$\begin{aligned} s &= \|(\Delta x, \Delta y)\|_2 \\ \mathbf{x}_{t+1} &= \mathbf{x}_t + s \begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} + \eta_{\text{motion}(0,1)} \\ \theta_{t+1} &= \theta_t + \Delta \theta + \eta_{\text{motion}(2)} \end{aligned} \quad (3)$$

The subscripts of η indicate the respective components of the (now) 3D noise vector, and Δ terms are odometry measurements. At test time for robot localization, position odometry is rotated to match the estimated heading before being passed to the map prior network to fix heading drift.

Particle weights are computed using the value of the location likelihood score heatmap at the particle’s location from the last n seconds of odometry. Low variance resampling of the particles is used due to its stability and consistent particle space coverage [22]. Resampling copies particles from the original cloud with probability proportional to their weight. The particle closest to the median (as opposed to the mean) of the new particle cloud is used as the agent’s estimated location. This guarantees the estimated location is not inside an obstacle even with a multi-modal particle distribution.

We also implement a re-initialization method for situations where the particle filter has diverged from a reasonable estimate. We reinitialize by sampling particles randomly within a radius r_{reinit} of the last estimated location when more than s_{reinit} of particles have passed through an obstacle. We experimentally set $r_{\text{reinit}} = 5$ meters and $s_{\text{reinit}} = 90\%$.

IV. DATASETS

We show the performance of our method on three datasets with varied sources of odometry.

IDOL Dataset: The IDOL dataset [8] is an inertial odometry dataset for indoor pedestrian localization. It consists of 20 hours of pedestrian trajectories collected by users carrying a smartphone in 3 buildings. Smartphone IMU data and ground truth device pose is available at 100 Hz. Odometry is generated from IMU data by the IDOL method proposed in the same work. 2D occupancy maps are generated from architectural plans available for the 3 buildings.

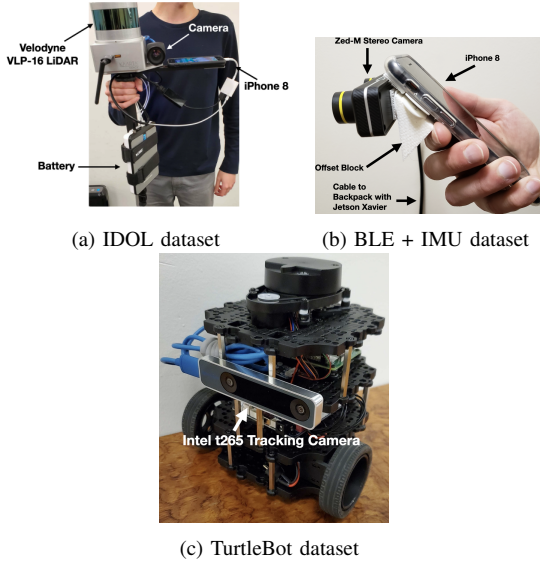


Fig. 4: Data collection setups used in our experiments.

BLE + IMU Dataset: We collect a smaller dataset similar to the IDOL dataset but with the addition of Bluetooth low energy (BLE) beacon measurements, to compare against bluetooth-based localization. Two floors of a building like those used in the IDOL dataset are instrumented with BLE beacons emitting a unique identifier broadcast at 1 Hz, which are recorded along with the broadcast signal strength and IMU data by a smartphone at 60 Hz. Ground truth device

pose and orientation are generated with a Stereolabs Zed Mini stereo camera configured as in Fig. 4b.

TurtleBot Dataset: A ROBOTIS TurtleBot 3 (Fig. 4c) was driven in 2 maps to collect 8 trajectories each of 10 minutes duration to evaluate generalization to wheel encoder odometry. Wheel odometry was generated using the onboard wheel encoder and IMU data, and ground truth pose was recorded from an attached Intel T265 tracking camera. The robot state (x, y, θ) follows a standard differential drive model:

$$\begin{aligned}\Delta s &= \pi r(n_L + n_R) \\ x_{t+1} &= x_t + \Delta s \cos(\Delta\theta + \theta_t) \\ y_{t+1} &= y_t + \Delta s \sin(\Delta\theta + \theta_t) \\ \theta_{t+1} &= \theta_t + \Delta\theta\end{aligned}\quad (4)$$

where r is the wheel radius, n_L and n_R are the number of revolutions recorded by the wheel encoders since the last time step, and $\Delta\theta$ is the relative orientation computed via a standard Madgwick filter [23] from IMU data.

V. EXPERIMENTS

We evaluate our method in three experimental settings.

(1) **IMU only** where we show the ability of our learned map prior to reduce drift in IMU-only odometry better than heuristic map priors. (2) **BLE + IMU** where we show that our IMU only method performs at a similar level to a BLE + IMU method that utilizes absolute positioning. (3) **Turtlebot** where we show the ability of our map prior network to generalize to unseen wheel encoder odometry.

Metrics: We evaluate localization performance using two main metrics: Absolute Trajectory Error (ATE) and End Error (EE). ATE is the root mean square error between points in the estimated and ground truth trajectories. The error at timestamp i is $e_i = \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2$. ATE measures global trajectory consistency and tends to increase with time due to drift. EE is the distance between the final estimated and final ground truth positions in a trajectory, i.e. $e = \|\mathbf{x}_T - \hat{\mathbf{x}}_T\|_2$. EE measures the total accumulated drift in the sequence.

Training/Testing: We implement the learnable map prior model in Pytorch Lightning [24] and train on a single Nvidia RTX 2080 Ti, with the Adam optimizer [25] and a learning rate of 0.01. A batch size of 32 is used and the network takes between 50 and 100 full passes through the data (epochs) to converge, depending on the map.

At test time the particle filter initializes with 1000 normally distributed particles ($\sigma = 0.01\text{m}$) around the true starting location. The filter runs at 1 Hz. For inertial odometry, the motion model uses white Gaussian noise with $\Sigma = 0.1\mathcal{I}_2$ m. $\Sigma = 0.01\mathcal{I}_3$ m is used for wheeled encoder odometry. A benefit of our formulation is the decomposition of trajectory and map processing. Once the network is trained, the map only needs to be processed once to obtain a Deep Map Tensor. This can be reused for all trajectories in the same map. At test time, only the trajectory LSTM needs to process data. An AMD Threadripper 1920x CPU takes 6.7 ms on average to compute the location likelihood score heatmap using a stored Deep Map Tensor. Our non-optimized particle

filter runs at 4x real-time speed, suggesting that our method would be feasible for smartphones or low-compute platforms. **IMU Only Experiments:** In our first experiment, we compare our method against four other localization methods to show the utility of our learned map prior. (1) Pedestrian Dead Reckoning (**PDR**) is implemented similarly to the baseline used in [6]. Step size is assumed to be 0.67 m, and the iOS internal estimate of device heading is used. (2) **IDOL** [8] is used to represent deep inertial odometry performance. We also compare against two approaches that use heuristic map priors. (3) One is a particle filter (**Heuristic PF**) implementation based on [2], where the prior is computed as the map cross correlation with a kernel based on noisy odometry. (4) The other is a linear-chain **CRF** implementation based on [3] that uses a hand-designed graph of possible user locations and transitions, and feature functions that capture the probability of transitions based on the observations. We use a unary feature function based on the absolute position and a pair-wise feature function based on relative odometry displacement. The feature function weights and graph edge length are chosen using grid search to achieve a good trade-off between accuracy and runtime.

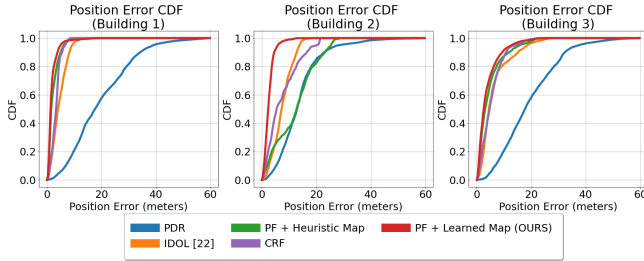


Fig. 5: Cumulative distribution function ATE error of our method and baselines from IDOL dataset experiments.

Tab. I shows the results of our localization performance on the IDOL dataset. Fig. 5 shows the error distributions. Our method significantly outperforms pure odometry and heuristic map-based methods across buildings in both ATE and EE. Fig. 1 shows the qualitative improvement in localization accuracy using our method. On average our method is able to achieve an error of less than 4 meters 80% of the time.

Fig. 1 shows that our method estimates more accurate locations and does not exhibit the large errors observed in heuristic methods. Notice that the heuristic PF often generates trajectories that penetrate walls similar to the odometry-only trajectory (Building 2). Building 2 has a long hallway which the odometry often drifts out of as users walk down it. Due to only small overlap of the trajectory with the walls, the heuristic prior weights the hallway-adjacent rooms similarly to the hallway. This leads to drift out of the hallway into these rooms. While the CRF trajectory does not pass through walls due to map constraints, it can erroneously enter rooms and ultimately become stuck. The CRF also outputs sharp unnatural trajectories because of the discretization of the map. Reducing the step resolution to generate a finer graph causes the inference to quickly become infeasible for real-time operation as the computation time grows quadratically

with graph size. Building 3 is substantially larger than buildings 1 and 2 and has narrow hallways. These factors lead to significant odometry drift which our method is able to handle consistently, except when the trajectory deviates too far away from the map which results in a slightly inferior EE accuracy.

Model	Bldg 1		Bldg 2		Bldg 3	
	ATE	EE	ATE	EE	ATE	EE
PDR	24.28	24.96	12.66	13.80	21.86	22.18
IDOL	5.65	8.51	6.62	10.61	8.33	9.71
CRF	4.66	6.58	7.56	10.54	8.82	7.38
Heuristic PF	3.11	3.42	11.37	15.36	6.71	9.67
Ours	2.87	1.60	2.51	6.08	5.66	9.99

TABLE I: Quantitative results of our approach compared to baselines on the IDOL [8] dataset. We report mean trajectory errors (in meters) over all sequences per building.

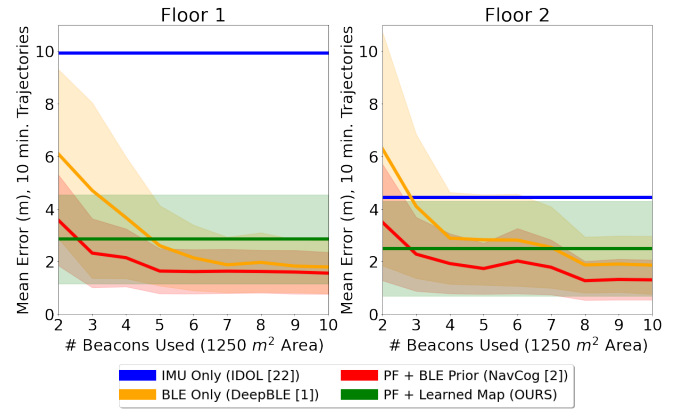


Fig. 6: Comparison of mean trajectory errors for varying bluetooth beacon densities for BLE + IMU experiments. Std. devs. shown by shaded regions. IDOL and our method do not use beacons so are independent of beacon density. Our method achieves accuracy similar to bluetooth-based absolute positioning without instrumentation of the environment.

BLE + IMU Experiments: We now compare our method using relative position measurements and a map to an absolute positioning method which directly estimates a user position in a global reference frame, *i.e.* BLE localization. BLE localization is implemented in two ways: (1) pure deep BLE localization (**BLE Only**) as in [9], and (2) using the deep BLE estimates in the particle filter observation model instead of a map (**PF + BLE Prior**) similar to NavCog [11].

Fig. 6 shows the results of using various numbers of beacons in a beacon localization system compared to IDOL and our method. BLE-only uses the raw, 1 Hz BLE estimates of location. PF + BLE Prior is able to achieve better performance by smoothing BLE estimates using inertial odometry (*i.e.* IDOL). Inertial odometry alone generally does not outperform BLE localization, but with the addition of map information through our learned prior we outperform BLE-only localization with fewer than 5-7 beacons. Overall, we achieve ATE within ~ 1 meter of both BLE methods using 8-10 beacons. Beacon densities of 8-10 have the best performance since many beacons can be used to precisely

triangulate the user. These experiments are performed in a $\sim 1250 \text{ m}^2$ area, which is fairly small relative to 10 beacons. Similar performance in larger areas would require hundreds of beacons which is not scalable (for both installation and maintenance). Our method is slightly less accurate but does not require instrumentation of the environment.

Model	Bldg 1		Bldg 2	
	ATE	EE	ATE	EE
Odometry	4.15	7.84	3.93	3.99
CRF	3.33	6.31	3.91	4.36
Heuristic PF	3.69	6.25	3.20	2.89
Ours	1.88	1.69	1.48	1.48

TABLE II: Quantitative results of our approach in comparison with baselines on the Turtlebot dataset. We report the mean trajectory errors (in meters) over all the sequences in a building.

Turtlebot Experiment: Tab. II reports results of our approach in tracking Turtlebot trajectories using wheel encoder odometry. Fig. 8 shows error distributions. Our learned prior in a particle filter halves localization error across buildings, which the heuristic methods are unable to handle. On average our method is able to achieve an error of 2.8 meters 90% of the time. This performance is achieved without retraining the network which has only ever seen inertial odometry.

The trajectories in Fig. 7 show our prior’s ability to accurately weigh particles. Odometry heading error often accumulates as the robot makes turns, but our learned prior absorbs some of this error and places weight in appropriate locations based on the general curving nature of the robot’s trajectory. The heuristic PF is unable to accurately weight regions when heading error occurs and so does not correct the error, leading to significant drift. The curvy nature of the trajectory also poses problems for the CRF as it is unable to capture these smooth transitions on the graph and instead estimates sharp lateral transitions which accumulate more error.

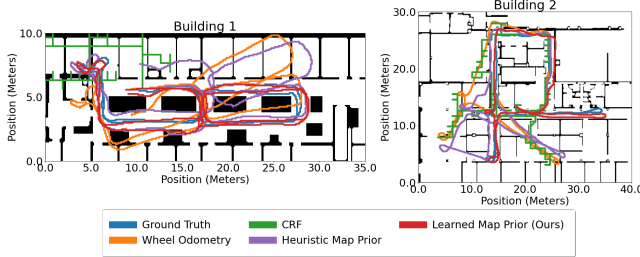


Fig. 7: Qualitative results showing the trajectories of the Turtlebot estimated by our method and other baselines.

Learned Map Prior Visualization: Fig. 9a shows a visualization of the map prior network output $\mathcal{S}(\mathbf{x}|M, O)$ for an input trajectory window. Spaces where the trajectory window input is feasible are generally given higher weight than extremely unlikely areas.

We also compare the KL divergence between a “ground

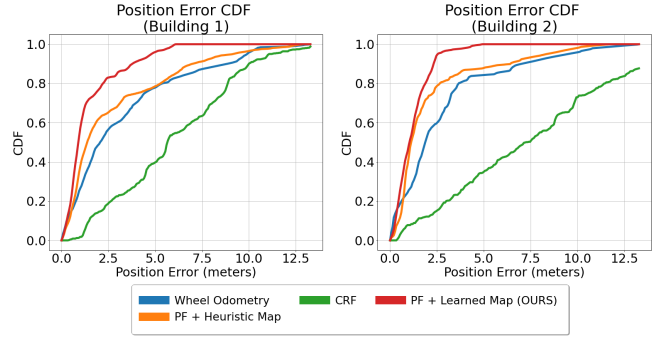


Fig. 8: Cumulative distribution function of ATE of our method and baselines from the Turtlebot experiments. Our method achieves best performance even without retraining on wheel odometry.

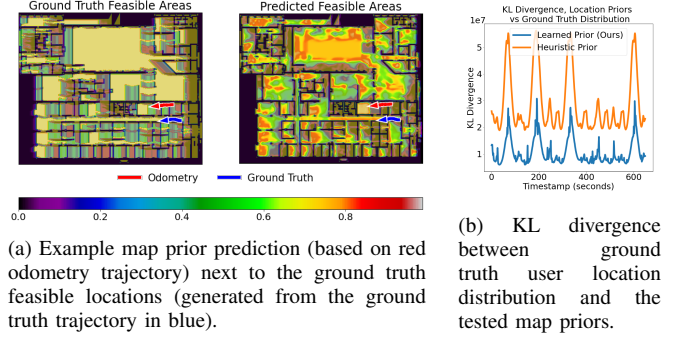


Fig. 9: Map prior visualization.

truth” location distribution G and our predicted prior distribution, and between G and the heuristic prior distribution, as shown in Fig. 9b. This indicates how well each prior places weight in the ground truth user location and avoids placing weight in unlikely locations. G is defined as a Gaussian ($\sigma = 1\text{m}$) around the ground truth user location. Our learned prior shows consistently lower divergence from the ground truth than the heuristic prior by a factor of approximately 3.

VI. CONCLUSION

In this work, we present a data-driven approach to combine occupancy map information with odometry measurements for indoor localization. Existing work uses hand-defined methods to incorporate map information. These methods make assumptions which generalize poorly or use noise-sensitive heuristics. Our method provides a more robust prior on indoor user location using a two-branch deep network utilizing separate learned spatial map and temporal odometry embeddings which are combined to determine feasible user locations in the map. Our prior, when used as a sensor model in a particle filter, is able to achieve 49% accuracy improvement over odometry for pedestrian localization. Compared to a BLE beacon-based localization system, our method exceeds BLE accuracy in low beacon-density spaces and approaches similar performance to BLE methods in high beacon-density spaces without requiring any devices installed in the environment. Our prior is also versatile, halving error in wheel encoder odometry-based localization for a robot without retraining. A key avenue of future work will be

addressing generalization of our map network to unseen maps through data collection in varied maps and synthetic data-generation techniques.

REFERENCES

- [1] J. Zhang and S. Singh, “Visual-lidar odometry and mapping: low-drift, robust, and fast,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. Seattle, WA, USA: IEEE, May 2015, pp. 2174–2181. [Online]. Available: <http://ieeexplore.ieee.org/document/7139486/>
- [2] A. Rechy Rormero, P. V. K. Borges, A. Pfrunder, and A. Elfes, “Map-Aware Particle Filter for Localization,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. Brisbane, QLD: IEEE, May 2018, pp. 2940–2947. [Online]. Available: <https://ieeexplore.ieee.org/document/8460707/>
- [3] Z. Xiao, H. Wen, A. Markham, and N. Trigoni, “Lightweight map matching for indoor localisation using conditional random fields,” in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, Apr. 2014, pp. 131–142.
- [4] R. A. MacLachlan and A. Dubrawski, “Applied indoor localization: Map-based, infrastructure-free, with divergence mitigation and smoothing,” *Inf. Fusion*, vol. 794, p. 801, 2010.
- [5] H. Xia, J. Zuo, S. Liu, and Y. Qiao, “Indoor Localization on Smartphones Using Built-In Sensors and Map Constraints,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 4, pp. 1189–1198, Apr. 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8444074/>
- [6] H. Yan, S. Herath, and Y. Furukawa, “RoNIN: Robust Neural Inertial Navigation in the Wild: Benchmark, Evaluations, and New Methods,” *arXiv:1905.12853 [cs]*, May 2019, arXiv: 1905.12853. [Online]. Available: <http://arxiv.org/abs/1905.12853>
- [7] W. Liu *et al.*, “TLIO: Tight Learned Inertial Odometry,” *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5653–5660, Oct. 2020, arXiv: 2007.01867. [Online]. Available: <http://arxiv.org/abs/2007.01867>
- [8] S. Sun, D. Melamed, and K. Kitani, “IDOL: Inertial Deep Orientation-Estimation and Localization,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 7, pp. 6128–6137, May 2021, section: AAAI Technical Track on Intelligent Robots. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/16763>
- [9] H. Agarwal, N. Sanghvi, V. Roy, and K. Kitani, “DeepBLE: Generalizing RSSI-based Localization Across Different Devices,” *arXiv:2103.00252 [cs]*, Feb. 2021, arXiv: 2103.00252. [Online]. Available: <http://arxiv.org/abs/2103.00252>
- [10] J. Zhang and S. Singh, “LOAM: Lidar Odometry and Mapping in Real-time,” in *Robotics: Science and Systems X*. Robotics: Science and Systems Foundation, Jul. 2014. [Online]. Available: <http://www.roboticsproceedings.org/rss10/p07.pdf>
- [11] D. Ahmetovic, C. Gleason, C. Ruan, K. Kitani, H. Takagi, and C. Asakawa, “NavCog: a navigational cognitive assistant for the blind,” in *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. Florence Italy: ACM, Sep. 2016, pp. 90–99. [Online]. Available: <https://dl.acm.org/doi/10.1145/2935334.2935361>
- [12] O. Pathak, P. Palaskar, R. Palkar, and M. Tawari, “Wi-fi indoor positioning system based on rssi measurements from wi-fi access points—a trilateration approach,” *International Journal of Scientific & Engineering Research*, vol. 5, no. 4, pp. 1234–1238, 2014.
- [13] S. Beauregard, Widyawan, and M. Klepal, “Indoor PDR performance enhancement using minimal map information and particle filters,” in *2008 IEEE/ION Position, Location and Navigation Symposium*. Monterey, CA, USA: IEEE, 2008, pp. 141–147. [Online]. Available: <http://ieeexplore.ieee.org/document/4570050/>
- [14] C. Chen, X. Lu, A. Markham, and N. Trigoni, “IONet: Learning to Cure the Curse of Drift in Inertial Odometry,” *arXiv:1802.02209 [cs]*, Jan. 2018, arXiv: 1802.02209. [Online]. Available: <http://arxiv.org/abs/1802.02209>
- [15] U. Onyekpe, V. Palade, A. Herath, S. Kanarachos, and M. E. Fitzpatrick, “WhOnet: Wheel Odometry Neural Network for Vehicular Localisation in GNSS-Deprived Environments,” p. 26.
- [16] A. Luo, S. Chen, and B. Xv, “Enhanced Map-Matching Algorithm with a Hidden Markov Model for Mobile Phone Positioning,” *ISPRS International Journal of Geo-Information*, vol. 6, no. 11, p. 327, Oct. 2017. [Online]. Available: <http://www.mdpi.com/2220-9964/6/11/327>
- [17] K. Zhao *et al.*, “DeepMM: Deep Learning Based Map Matching with Data Augmentation,” in *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Chicago IL USA: ACM, Nov. 2019, pp. 452–455. [Online]. Available: <https://dl.acm.org/doi/10.1145/3347146.3359090>
- [18] D. Simon, *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. John Wiley & Sons, 2006.
- [19] J. S. Liu and R. Chen, “Sequential Monte Carlo Methods for Dynamic Systems,” *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, Sep. 1998. [Online]. Available: <http://www.tandfonline.com/doi/full/10.1080/01621459.1998.10473765>
- [20] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *arXiv:1505.04597 [cs]*, May 2015, arXiv: 1505.04597. [Online]. Available: <http://arxiv.org/abs/1505.04597>
- [21] S. Padhy, “shreyaspadhy/U-Net-Zoo,” May 2021, original-date: 2017-12-09T20:48:15Z. [Online]. Available: <https://github.com/shreyaspadhy/U-Net-Zoo>
- [22] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*, ser. Intelligent robotics and autonomous agents. Cambridge, Mass: MIT Press, 2005, oCLC: ocm58451645.
- [23] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, “Estimation of IMU and MARG orientation using a gradient descent algorithm,” in *2011 IEEE International Conference on Rehabilitation Robotics*. Zurich: IEEE, Jun. 2011, pp. 1–7. [Online]. Available: <http://ieeexplore.ieee.org/document/5975346/>
- [24] W. Falcon *et al.*, “PyTorchLightning/pytorch-lightning: 0.7.6 release,” May 2020. [Online]. Available: <https://github.com/PyTorchLightning/pytorch-lightning>
- [25] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980. [Online]. Available: <http://arxiv.org/abs/1412.6980>