

Trajectory Imputation Using Computer Vision Models

Panagiotis Betchavas* Alexandros Troupiotis-Kapeliaris^{†*} Kostas Patroumpas*
Giannis Spiliopoulos^{†*} Dimitrios Skoutas[‡] Dimitris Zisis[†] Nikos Bikakis^{§*}
*Archimedes/Athena RC, Greece [†]University of the Aegean, Greece [‡]Athena RC, Greece
[§]Hellenic Mediterranean University, Greece

Abstract—The completeness of moving object trajectory data is critical for accurate transportation monitoring and analysis, as well as for enabling reliable downstream analytics. In this work, we propose IMGIN, a novel framework that reformulates trajectory reconstruction as an image inpainting task. Our framework introduces the *wave map*, a spatiotemporal representation that maps raw trajectory data onto a structured grid via a universal indexing system, thereby enabling the use of high-performance computer vision architectures for path recovery. Preliminary evaluation on real-world maritime datasets demonstrates that IMGIN achieves accuracy comparable to the state-of-the-art while providing improved spatial interpretability.

Index Terms—Image inpainting, Image-based trajectory reconstruction, U-Net, CNN, Spatiotemporal encoding, Rasterized trajectories, Visual representation learning, Mobility analytics

I. INTRODUCTION

Reliable tracking of moving objects is a cornerstone of Intelligent Transportation Systems, with modern protocols providing high-precision, near real-time positioning. However, recorded trajectories often contain missing segments due to coverage gaps, hardware failures, or signal interference. For example, in the maritime domain, where the Automatic Identification System (AIS) generates massive volumes of high-frequency historical data, these gaps present a significant challenge for global vessel activity monitoring [1]. This data abundance allows modern deep learning approaches to learn underlying mobility patterns, significantly improving the fidelity of trajectory imputation [2]. Furthermore, a direct parallel can be drawn between this recovery process and the task of image inpainting. A variety of computer vision models have already demonstrated remarkable accuracy in high-stakes reconstruction tasks [3], like detail restoration in medical imaging or temporal continuity in video streams. Despite their proven ability to recover complex structures, the application of these powerful architectures to the spatial recovery of trajectory data remains a largely underexplored domain.

To this end, we introduce IMGIN, a framework for IMaGe-based ImputationN of Trajectories. Our approach utilizes the H3 indexing system to transform trajectory data into structured *wave maps*, which are then processed through a Convolutional Neural Network (CNN) architecture to reconstruct missing segments via image-infilling. Without loss of generality, this work currently focuses on a maritime context, but the proposed approach can be applied to other mobility settings, such as urban traffic or robotic navigation. We provide a publicly available implementation integrating standard Python libraries to facilitate further research in maritime analytics¹.

¹Source code and implementation: <https://github.com/M3-Archimedes/IMaGe-based-ImputationN-of-Trajectories>

Related Work. Vessel trajectory imputation is closely related to path prediction and has been addressed through diverse methodologies [2]. Beyond classical geometric approaches or data source fusion [4], data-driven methods facilitate the integration of local movement patterns by analyzing historical AIS data. These often utilize movement networks, such as graphs or common route clusters, to infer missing segments [1]. In a multi-scale solution, the multi-scale heterogeneous graph constructed in [5] can capture correlations between AIS observations and features for trajectory imputation. Another prominent category leverages deep learning architectures to model trajectories as spatiotemporal sequences. These approaches typically prioritize the temporal dimension, utilizing Recurrent Neural Networks like GRUs and LSTMs to capture sequential dependencies [6], [7]. Recent shifts toward generative models include the use of Large Language Models (LLMs) for sequential reasoning [8] and diffusion techniques for noise-based reconstruction [9]. Notably, a physics-guided diffusion model is proposed in [10] for long-term vessel trajectory imputation, incorporating vessel mobility constraints to enforce physically consistent outputs. While these approaches focus on improving the modeling of trajectory dynamics and missing data reconstruction, they still operate primarily in sequence or graph-based representations. Visual analytics in the maritime domain are mostly focused on either traffic density representations or the classification of vessel activity [11], [12]. To the best of our knowledge, vessel trajectory imputation has not been explicitly formulated as an image reconstruction problem over rasterized trajectory representations.

Contributions. Here is a summary of our main contributions:

- We reframe the trajectory imputation problem as an image-based reconstruction task. This novel perspective allows the problem to benefit from advanced computer vision techniques for spatial continuity.
- We define a novel *wave map* representation, mapping raw trajectory data into structured, multi-channel images. This formulation preserves movement dynamics within a discrete global grid.
- We introduce IMGIN, a complete open-source pipeline for trajectory image modeling, featuring a CNN-based architecture specifically optimized to fill trajectory gaps.
- We conduct a preliminary evaluation using real-world maritime data, highlighting the potential of our approach and demonstrating that it effectively competes with a state-of-the-art baseline.

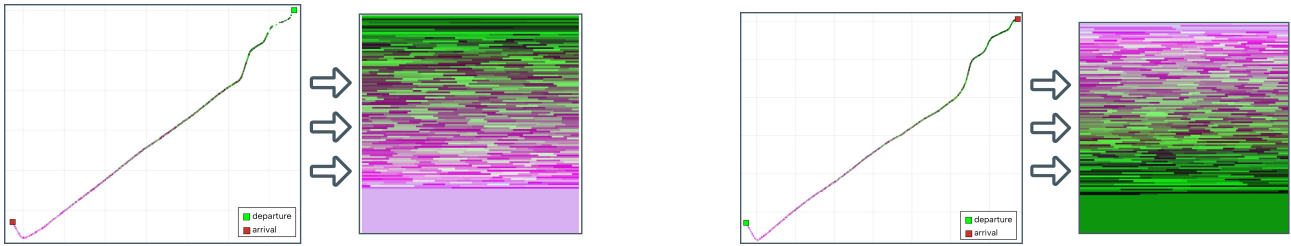


Fig. 1: Transformation of two full trajectories of opposite directions to their equivalent *wave maps*.

II. METHODOLOGY

The proposed framework treats trajectory gap reconstruction as an image inpainting problem [3]. Instead of predicting missing longitude/latitude values directly in sequential form, each trajectory is transformed into a structured *wave map* in which temporal evolution is represented by image coordinates and spatial location is encoded by color. Under this formulation, missing trajectory segments appear as masked image regions, and reconstruction is achieved by recovering the missing color patterns from the surrounding context. This representation enables the use of convolutional encoder–decoder architectures while preserving the joint temporal and spatial structure. After reconstruction in image space, the result is mapped back to the geographic domain to yield a completed trajectory.

A. Trajectory Data Preprocessing

Filtering and Trip Segmentation. Given a period of interest, we isolate trajectories traveling between a specific pair of ports, partitioning the data into discrete trips based on unique identifiers and port locations. To ensure temporal consistency for modeling, we exclude trips with outlier durations, such as brief in-harbor movements or extended loitering. This process ensures the final dataset consists of standard transits along a consistent route, with uniform durations.

Temporal Augmentation. Because AIS observations are irregular in time, trajectories are first resampled onto a regular temporal support. Each trip is projected onto N temporal bins defined over a common duration equal to the longest trip duration observed in the dataset. When multiple observations fall within the same bin, the observation closest to the bin center is selected as the representative sample. Empty bins are then filled using interpolation from neighboring bins.

H3-Based Spatial Encoding. To establish a constrained spatial vocabulary and simplify the representational space, we discretize the positions using the H3 hexagonal indexing system² at a constant resolution. This process first maps the trajectories into discrete H3 sequences, effectively quantizing the continuous geographic space. To ensure numeric continuity between neighboring cells, we represent each cell by its centroid, calculated as the mean longitude and latitude of all trajectory points assigned to that cell. As a final step, these

centroids are normalized relative to the geographic bounding box of the dataset to facilitate consistent bit-packed encoding.

B. Visual Representation

RGB Mapping. To transform raw trajectories into a computer-vision-compatible representation, we encode each normalized H3 centroid as an RGB triplet. Channel values are stored as float32 numbers, and each normalized latitude and longitude ordinate is quantized to 23 bits (mantissa precision). The red channel captures latitude’s high 18 bits, the blue channel longitude’s high 18 bits, and the green channel packs together the two 5-bit remainders. The resulting channel values are inverted and rescaled to the interval $[0.05, 0.95]$, thereby avoiding pure black and pure white values reserved for visualization extremes and masking. This encoding yields a unique vocabulary over the observed H3 cells while remaining compatible with float32 tensor storage and approximate reverse decoding.

Because each pixel corresponds to the color of a valid H3 cell rather than to unconstrained continuous coordinates, reconstruction is naturally constrained to a discrete spatial vocabulary. In decoding, reconstructed colors are mapped back to geographic space either through nearest H3 centroid lookup or through approximate inversion of the bit-packed representation. In this way, the model predicts within a structured set of observed spatial states, which improves interpretability and helps keep reconstructed trajectories geographically consistent.

Wave-map Generation. At the trip level, each augmented trajectory is further transformed into a $N \times N$ “*wave map*” tensor, as illustrated in Figure 1. In this representation, the vertical axis consists of the N sequential temporal bins of a fixed duration, progressing downward from the first row. The horizontal axis consists of N uniformly spaced temporal sampling positions within each bin. For a given row, the method identifies the observation whose timestamp is closest to each horizontal sampling position. Consequently, temporal structure is encoded through the tensor’s grid position, while the specific spatial data is stored within the pixels.

Overall, this strategy transforms continuous movement into a structured visual language, regularizing reconstruction through a discrete spatial vocabulary while allowing for precise coordinate recovery via nearest-neighbor centroid lookup.

²<https://h3geo.org>

C. H3-Aware U-Net Architecture

Architecture. After the map generation, the modeling and imputation are achieved through a U-Net encoder–decoder architecture [13] operating on four-channel inputs. The first three channels correspond to the partially observed RGB wave map, and the fourth channel is a binary mask indicating missing regions. The encoder comprises four convolutional stages, each containing two 3×3 convolutions with batch normalization and ReLU activations, followed by max-pooling. A bottleneck layer with dropout provides regularization and enlarges the effective receptive field. The decoder mirrors the encoder by means of transposed convolutions and skip connections, enabling the model to reconstruct both large-scale structure and local detail. A final 1×1 convolution followed by a sigmoid activation produces the three-channel RGB output.

To ensure spatial validity of the reconstructed image, predicted RGB values are quantized to the nearest valid H3 color by means of a Kd-tree over the H3 color vocabulary. This guarantees that reconstructed pixels correspond to admissible H3 cells and eliminates geometrically inconsistent color outputs.

Training Protocol. Training samples are created by masking contiguous horizontal bands of the wave map at multiple scales. Masks span up to 30% of the image height, excluding the first and last few rows to preserve boundary context. Optimization is performed using AdamW with a OneCycle learning-rate schedule until convergence. The model is trained using a weighted combination of Mean Squared Error and Mean Absolute Error, specifically optimized to prioritize the reconstruction of missing regions while maintaining global trajectory consistency across the full wave map.

D. Trajectory Reconstruction

During inference, significant (non-trivial) missing intervals are filled through a multistep process. First, the full trajectory is processed by the trained model, filling the missing (masked) image segments. The reconstructed image region is then quantized using a Kd-tree structure to translate RGB vectors into valid H3 colors. The image is then sampled at regular temporal intervals (e.g., 2 minutes); for each timestamp, the corresponding wave map row is examined to map high-confidence pixels to their H3 cell centroids. In cases of multiple cell predictions, a single median spatial estimate is calculated for each timestamp.

These image-derived positions serve as internal anchor points within the missing interval. Final reconstruction is obtained by interpolating between known boundary observations and these anchor points to fill the remaining timestamps. By mapping the model’s spatial predictions directly to the image’s vertical temporal grid, the solution ensures that every reconstructed position is explicitly tied to a precise chronological coordinate, resulting in more complete outcomes.

III. EXPERIMENTAL EVALUATION

A. Experimental Setup

Dataset. In order to evaluate the accuracy of our framework, we conducted preliminary experiments on a real-world AIS dataset

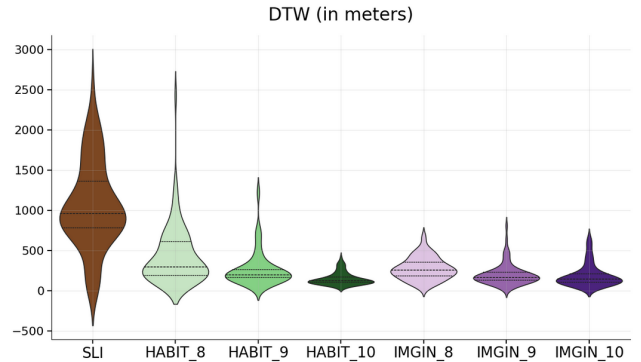


Fig. 2: DTW error distributions for IMGIN and HABIT across H3 resolutions 8–10, compared against grid-independent SLI.

sourced from the registry of the Danish Maritime Authority. More precisely, we focused on the highly traversed passenger route between the ports of Frederikshavn in Denmark and Gothenburg in Sweden for a three-month period (Jan–March 2024). The processed dataset comprises 577,940 AIS positional messages from 267 complete trips across both transit directions. We set $N = 128$ for temporal augmentation and wave-map generation, resulting in images of 128×128 pixels. For training, we utilized 70% of the total trips, reserving the remaining 81 voyages for evaluation. To assess reconstruction performance, we introduced synthetic gaps of varying durations (25–65 min) and compared the resulting imputations against the ground truth trajectories.

Methods. We evaluate the proposed framework against two methods. The first, HABIT [1], is a recently published approach that similarly leverages the H3 indexing system, enabling a direct comparison across identical spatial resolutions (8–10). HABIT constructs a local graph based on historical transitions to perform trajectory imputation. Additionally, we include Straight-Line Interpolation (SLI) as a naive baseline, which assumes a constant course and speed, connecting gap endpoints via a direct linear path.

Metrics. We use two metrics to evaluate reconstruction fidelity: (1) Dynamic Time Warping (DTW), which measures cumulative spatial distance between sequences, and (2) Longest Common Subsequence (LCSS), which quantifies the percentage of points within a 250m threshold. This threshold represents the approximate length of a large passenger vessel to ensure physical plausibility.

B. Results and Discussion

As illustrated in Figure 2, both IMGIN and HABIT offer substantial improvements over the SLI baseline, with performance scales with spatial refinement. Specifically, IMGIN outperforms HABIT across resolutions 8 and 9 in both mean and median DTW scores (Table I). At resolution 10, the performance gains of IMGIN narrow, likely due to the increased H3 vocabulary size requiring more extensive training data to fully capture the higher-dimensional spatial transitions. Nevertheless, IMGIN

H3	DTW (meters)				LCSS ($\epsilon=250m$)			
	HABIT		IMGIN		HABIT		IMGIN	
	Mean	50%	Mean	50%	Mean	50%	Mean	50%
8	435.7	296.2	277.0	254.9	0.17	0.18	0.36	0.35
9	271.1	198.5	201.9	167.7	0.16	0.16	0.48	0.49
10	150.3	123.0	190.7	142.8	0.21	0.21	0.50	0.52

TABLE I: Comparative reconstruction performance of IMGIN and HABIT across H3 resolutions for DTW and LCSS.

maintains average DTW errors below 300m (comparable to the length of a large passenger vessel) demonstrating its practical utility in maritime setting.

Prediction accuracy is also correlated with local traffic constraints, as seen in Figure 3. In open waters, the broader distribution of bidirectional trips creates a significantly larger set of viable H3 cells; this increased search space offers more alternative paths, which can lead to higher error rates. Conversely, performance improves in confined waters, such as straits or port approaches, where restricted movement options naturally limit the model’s output space. Moreover, IMGIN’s LCSS superiority stems from the bit-packed H3 representation, which transforms discrete cell IDs into a spatially continuous coordinate system where color proximity reflects physical distance. This allows the U-Net to produce a distributed wave map of probability that, when averaged during reconstruction, effectively keeps the predicted path within the 250m threshold.

IV. CHALLENGES AND FUTURE WORK

In this work, we present a framework for trajectory imputation that is based on computer-vision techniques. Representing trajectories as *wave maps* enables the training of a CNN-based model that achieves reconstruction performance comparable to the current state-of-the-art. A few challenges have been identified in this process, and should be addressed in future work. More prominently, this includes the creation of a model suitable for trips between different stops (ports) and by vessels of different types (passenger, cargo, fishing). Beyond the maritime domain, this framework could be applied to other mobility settings, such as urban vehicle traffic, to test its generalizability across diverse movement patterns. A key component of this generalization will be the effective modeling of variant trip durations, potentially addressed through adaptive temporal scaling or by parsing trajectories into localized segments managed by specialized sub-models. Additionally, the inclusion of voyage-specific features, like weather conditions or activity type, may be incorporated as values in columns like the color opacity to enhance the accuracy of such approaches. Furthermore, accuracy could be enhanced by incorporating exogenous voyage-specific features, such as weather conditions, represented via additional data channels. On a broader scale, developing models that capture the entirety of regional traffic could pave the way for large-scale pretrained foundation models, or be integrated with the reasoning capabilities of Large Language Models (LLMs) via textual trajectory representations

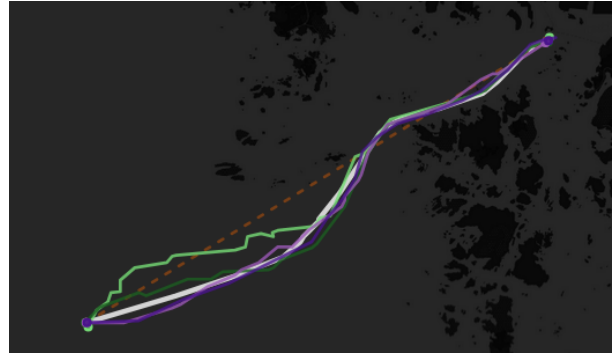


Fig. 3: Imputation comparison: ground truth (white), SLI (brown), HABIT (light/dark green for H3 resolution 9/10), and IMGIN (pink/purple for H3 resolution 9/10).

[14], leading toward specialized maritime Vision-Language Models (VLMs).

Acknowledgments. This work has been partially supported by project MIS 5154714 of the National Recovery and Resilience Plan Greece 2.0 funded by the European Union under the NextGenerationEU Program.

REFERENCES

- [1] G. Spiliopoulos, A. Troupiotis-Kapeliaris, K. Patroumpas, N. Liapis, D. Skoutas, D. Zissis, and N. Bikakis, “Data-Driven Trajectory Imputation for Vessel Mobility Analysis,” in *EDBT*, 2026.
- [2] A. Graser, A. Jalali, J. Lampert, A. Weissenfeld, and K. Janowicz, “Mobilitydl: a review of deep learning from trajectory data,” *Geoinformatica*, vol. 29, no. 1, 2025.
- [3] W. Quan, J. Chen, Y. Liu, D. Yan, and P. Wonka, “Deep learning-based image and video inpainting: A survey,” *Int. J. Comput. Vis.*, vol. 132, no. 7, 2024.
- [4] S. Wu, K. Torp, A. Troupiotis-Kapeliaris, D. Zissis, E. Zimányi, and M. Sakr, “Effective ship trajectory imputation with multiple coastal cameras,” in *IEEE MDM*, 2025.
- [5] H. Liu, T. Li, Y. He, K. Torp, Y. Li, and C. S. Jensen, “Mh-gin: Multi-scale heterogeneous graph-based imputation network for ais data,” *PVLDB*, 2025.
- [6] M. Liang, R. W. Liu, Q. Zhong, J. Liu, and J. Zhang, “Neural network-based automatic reconstruction of missing vessel trajectory data,” in *IEEE ICBDA*, 2019.
- [7] J. Chen, M. Liang, C. Peng, J. Zhang, and S. Huo, “Improving maritime data: A machine learning-based model for missing vessel trajectories reconstruction,” *IEEE Transactions on Vehicular Technology*, 2025.
- [8] H. Liu, T. Li, H. Wang, K. Torp, T. Zhang, Y. Li, and C. S. Jensen, “Vista: Knowledge-driven interpretable vessel trajectory imputation via large language models,” *arXiv:2601.06940*, 2026.
- [9] W. Wang, W. Xiong, L. Chen, and H. Chen, “Trajdiff: A method for vessel trajectory imputation utilizing resampled conditional diffusion models,” *Results in Engineering*, 2025.
- [10] Z. Zhang, Z. Fan, Z. Lv, X. Song, and R. Shibusaki, “Long-term vessel trajectory imputation with physics-guided diffusion probabilistic model,” in *ACM KDD*, 2024.
- [11] I. Kontopoulos, A. Makris, D. Zissis, and K. Tserpes, “A computer vision approach for trajectory classification,” in *IEEE MDM*, 2021.
- [12] E. d’Afflisio, L. M. Millefiori, P. Braca, and M. Guerriero, “Maritrac: Maritime trajectory classification using object instance segmentation with model-based generated data augmentation,” in *International Conference on Information Fusion (FUSION)*, 2024.
- [13] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Int. Conf. on Medical image computing and computer-assisted intervention*, 2015.
- [14] K. Patroumpas, A. Troupiotis-Kapeliaris, G. Spiliopoulos, P. Betchavas, D. Skoutas, D. Zissis, and N. Bikakis, “Context-enriched natural language descriptions of vessel trajectories,” in *IEEE MDM*, 2026.