Robotics: Science and Systems 2026
Sydney, Australia, July 13–17, 2026

AnyAmber: A Generalist for Versatile Anonymous Bearing and Range Based Position Tracking

Si Wang*, Yanmei Jiao†, Yanjun Cao‡, Rong Xiong*, Yue Wang*§
*Institute of Cyber-Systems and Control, Zhejiang University, China
†School of Information Science and Engineering, Hangzhou Normal University, China
‡Huzhou Institute of Zhejiang University, Huzhou, China
§Corresponding author: ywang24@zju.edu.cn

***Abstract*—Position tracking based on bearing measurements and Ultra-wideband (UWB) ranging is widely used in robotic navigation tasks. However, due to variations in the number of robots, anchor configurations, UWB tag layouts, and the presence or absence of anonymous visual observations, existing methods typically rely on specially designed solvers or networks tailored to a particular localization problem. In this work, we introduce AnyAmber, a generalist neural network for versatile anonymous bearing and range based position tracking. To model diverse and dynamic geometric constraints, a heterogeneous EGAT network is employed for unified localization and uncertainty estimation, cascaded with a differentiable hierarchical PGO for improved accuracy. Additionally, we incorporate a Sensor-Embedded GRU module for adaptive UWB bias correction and a temporal graph-based matching network for soft assignments between robots and anonymous bearings. By adopting a unified problem formulation, our model is jointly pretrained on a large-scale multi-task dataset encompassing diverse simulated and real-world environments. In the experiments, with fine-tuning on only a single trajectory in a target test scenario, the model achieves superior few-shot localization performance to existing methods. Our code is available at: https://github.com/HiOnes/AnyAmber.**

I. Introduction

Precise position tracking serves as a fundamental prerequisite for both single robot and robot swarms to carry out complex tasks such as exploration, search and rescue. In scenarios without access to global positioning aids such as GPS [1–3] or motion capture systems [4], position tracking must rely entirely on relative measurements. Bearing and range observations constitute the core modalities for establishing spatial relationships between agents and their environment. In single-robot localization applications, range measurements between multiple fixed anchors and the robot-mounted UWB tags are utilized to estimate the robot's position with respect to the map reference frame. In multi-robot systems, pairwise distance measurements and anonymous visual bearings serve as constraints for estimating their relative poses in the local frame. Despite the shared geometric foundations, as shown in Fig. 1, these problems are often treated as distinct challenges depending on the configuration: single vs. multi-robot, or single vs. multi-tag.

For various tasks, existing approaches [5–22] have focused on optimizing performance on individual problems by designing specialized solvers or networks. Although task-specific neural networks [6–14] can achieve state-of-the-art performance under particular localization settings, they lack versatility, a network designed for single-robot localization cannot easily adapt to multi-agent relative tracking. Conversely, general frameworks like Pose Graph Optimization (PGO) offer the necessary flexibility to model various configurations but often fail to match the accuracy of learning based methods when faced with non-Gaussian noise, signal multipath, or complex disturbances. Consequently, a *unified* and *learnable* solution capable of addressing diverse localization problems remains an open challenge.

Recent studies in visual geometry and semantics [23–26]

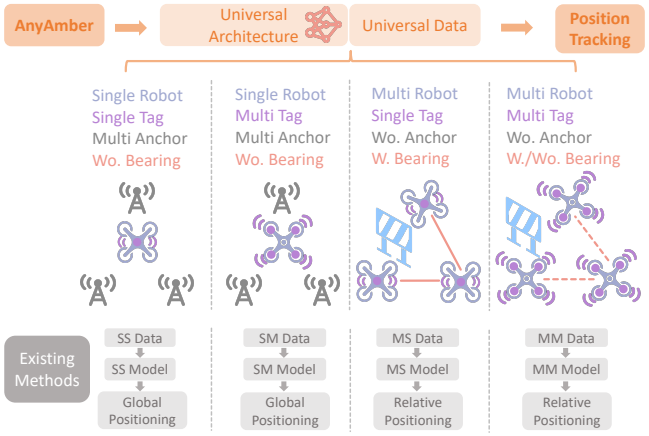


Fig. 1: Overview of our universal neural localization approach.

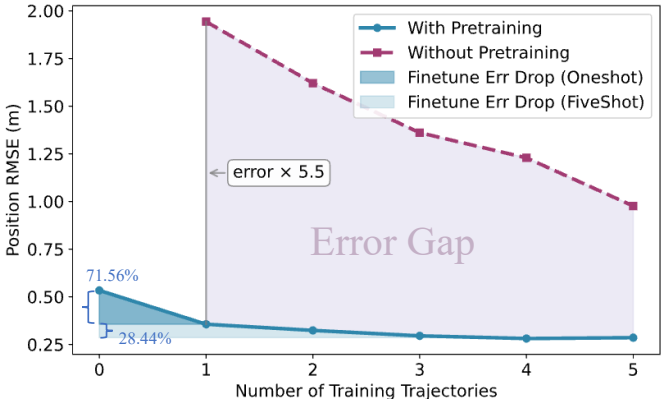


Fig. 2: Localization accuracy vs. number of training trajectories for pretrained and non-pretrained models.

reveal that universal network architectures not only enhance model usability but also improve single-task performance through learning the generalizable features across multiple tasks. Inspired by these findings, we investigate whether the generalizable knowledge can be captured for enhanced performance through joint pretraining in the context of bearing and range based position tracking.

To bridge the fragmentation while simultaneously boosting performance on individual tasks, in this paper, we propose **AnyAmber**, a generalist network to handle versatile anonymous bearing and range based position tracking tasks, covering single-robot single-tag (SS), single-robot multi-tag (SM), multi-robot single-tag (MS), and multi-robot multi-tag (MM) scenarios, which are illustrated in Fig. 1. To achieve the above objectives, we require two key components: a unified problem formulation and high-fidelity measurements (assigned bearings, low-noise UWB rangings). Specifically, we bridge the gap between single- and multi-robot positioning by reformulating them into a unified relative localization problem. Built upon this consistent formulation, a heterogeneous EGAT network is utilized to aggregate multi-sensor geometric features, yielding an accurate estimation of the robot's pose and covariance. For data association between anonymous visual observations and robots, we design a temporal graph attentional neural network in conjunction with a multi-frame Sinkhorn algorithm, ensuring accurate and reliable matching. To compensate for UWB signal disturbances, a Sensor-Embedded GRU network is employed for adaptive ranging bias correction. Moreover, to improve overall accuracy while enabling end-to-end training, we incorporate a two-stage differentiable PGO scheme.

Empowered by the universal architecture, our model gains transferable localization capability through joint pretraining on a large-scale unified dataset, covering 108 diverse scenarios, 225 trajectories and 153555 sampled frames. Comprehensive experiments verify that with only a single trajectory for fine-tuning in a specific unseen environment, the proposed AnyAmber outperforms the existing specialized models, significantly reducing the demand for data collection and boosting performance on individual tasks, as few-shot performance presented in Fig. 2. Overall, our major contributions are as follows:

- Unified formulation covering various localization tasks: single- and multi-robot settings, single- and multi-tag configurations, with or without anchor setups, and bearing or non-bearing measurements.
- Generalist graph attention network **AnyAmber** for training and inference of versatile anonymous bearing and range based position tracking tasks.
- Modules to improve the performance, including GRU for UWB ranging denoising, temporal GNN-based match network for soft data association, and a hierarchical differentiable PGO structure for optimization.
- Substantial experimental results validating the cross-task training enables superior performance in previously unseen environments over task-specific methods, requiring only few-shot fine-tuning.

II. Related Work

In this section, we introduce several task-specific position tracking architectures based on visual and range measurements, as well as a set of universal multi-task models.

A. Specialized Position Tracking Architecture

1) Single-robot Positioning: Single-robot localization aims to recover the robot's position in a global frame from UWB tag-to-anchor ranges, possibly fused with IMU measurements. Current solutions generally fall into four classes: extended Kalman filter (EKF) frameworks [27, 28], nonlinear optimization formulations [5, 29], B-spline trajectory representations [30, 31], and learning-based models [6–13]. EKF-based methods [27, 28] fuse ranging and inertial measurements via the Kalman filtering framework, while the loosely-coupled integration scheme and Gaussian noise assumption may result in suboptimal localization performance. Nonlinear optimization–based approaches [5, 29] can effectively integrate heterogeneous sensing modalities. Nonetheless, these methods are sensitive to the quality of initial conditions. Poor initialization or significant range errors induced by non-line-of-sight (NLOS) can cause the optimizer to fail to converge or to become trapped in local minima. B-spline approaches [30, 31] support interpolation at any time instant while enforcing continuous motion-model constraints, making them well suited for scenarios with high ranging noise or sparse observations. Yet, they also exhibit strong sensitivity to initialization, and necessitate extensive hyperparameter tuning such as knot density and sliding window size. Learned methods [6–13] are capable of modeling the non-Gaussian distribution of UWB ranging errors and exploiting multimodal information in a tightly-coupled manner, leading to improved localization performance. VIUNet [7] features a CNN-LSTM module for visual-inertial odometry and a MLP module for UWB absolute localization, respectively, achieving visual-inertial-ranging fusion for single-robot single-tag localization. MLP-, CNN-, and LSTM-based [6, 7] networks struggle to handle dynamic UWB nodes, whereas graph neural network (GNN)-based methods [9–13] stand out for UWB localization due to their dynamic topological structure and superior feature aggregation capabilities. NRIO [13] models the UWB localization subsystem using a graph neural network and couples it with GRU-driven inertial odometry, enabling robust, high-accuracy multi-tag localization on a single robot.

2) Multi-robot Positioning: In a multi-robot localization system, each robot estimates the positions of its neighboring robots in its own local coordinate frame. Due to the underdetermined nature of the multi-robot relative localization problem, inter-robot distance measurements alone are typically insufficient to maintain observability, necessitating auxiliary bearing information from Angle-of-Arrival (AoA) [15, 16] or vision-based methods [14, 17–22]. Visual detections deliver explicit and highly informative bearing observations but lack identity labels by nature, rendering the resulting data-association problem particularly challenging. Consequently, prior work has investigated both hardware-encoded schemes

[17–19] and algorithmic matching-based solutions [14, 20–22] to robustly resolve the measurement correspondence ambiguity. In this work, we focus on the most challenging scenario, realizing multi-robot relative localization relying solely on anonymous visual detections and inter-robot ranging. As a representative approach within this line of research, Mr. Virgil [14] adopts GNN and Sinkhorn algorithm for data association, with a learnable match front-end and a differentiable PGO back-end, achieving robust multi-robot tracking. Notably, in multi-robot configurations equipped with multiple antennas per robot, explicit bearing observations are optional. MURP [32] develops a multi-tag relative localization system, achieving low communication cost with pure range measurements.

B. Unified Architecture

Methods built upon the above specialized architectures can achieve excellent performance for their designated localization problems. However, their task-specific problem formulations and network designs limit their applicability and transferability to broader localization settings.

Within the image segmentation literature, extensive studies [23–25] have explored unified frameworks that jointly address semantic, instance, and panoptic segmentation, and such universal models have demonstrated superior performance compared with specialized architectures on individual segmentation tasks. Mask2Former [24] requires individual training on specific tasks to achieve optimal performance, while OneFormer [25] constructs a unified dataset to jointly train a single and universal model. In the field of simultaneous localization and mapping (SLAM), SLAM-Former [33] unifies the visual SLAM workflow (tracking and incremental mapping front-end, optimization back-end) within one Transformer, achieving online dense monocular SLAM with global consistency.

However, within the domain of position tracking based on anonymous bearing and ranging information, no existing approach provides a unified solution capable of addressing the full spectrum of scenarios, including both single and multiple robots as well as single-tag and multi-tag configurations.

III. Formulation of Unified Position Tracking

In this section, we begin by defining the notations and symbols used throughout this paper. Subsequently, we discuss the most complex but general localization case, namely the multi-robot multi-tag localization task. Finally, we demonstrate how diverse position tracking tasks can be embedded into a unified formulation by expressing them as special cases of the general multi-robot multi-tag localization problem.

A. Notation

We denote the world (or map) coordinate frame as W, the static anchor as A, the robot base frame as B, and the tag attached to the robot as T. The position and orientation of the robot base frame with respect to the world frame are written as $\mathbf{t}_B^W \in \mathbb{R}^3$ and $\mathbf{R}_B^W \in \mathbb{R}^{3\times3}$ respectively, where the rotation matrix is parameterized by quaternion in our implementation. The 6-DoF pose of frame B relative to the frame W is denoted by $\mathbf{P}_B^W$, which is equivalent to the transformation matrix $\begin{bmatrix}\mathbf{R}_B^W & \mathbf{t}_B^W \\ 0 & \mathbf{1}\end{bmatrix}$. The estimated quantities are represented by $\hat{(\cdot)}$. The symbol $\|\cdot\|$ denotes the Euclidean norm of a vector. Let N_A, N_R, N_T, N_C, and N_W denote the numbers of UWB anchors, robots, UWB tags per robot, max camera bearing detections, and time window size in multi-frame matching, respectively. Considering the possibility of erroneous observations, N_C is set larger than $N_R - 1$. $\mathcal{S}_\mathcal{R} = \{R_i | 1 \le i \le N_R\}$ is the set of robot swarms.

TABLE I: Unified multi-robot multi-tag (MM) formulation for four typical positioning tasks. # R, # T, # A refer to the number of robots, tags per robot, and anchors, respectively.

Task		Original Setting			Unified MM Formulation			
					Reference		Estimated	
		# R	# T	# A	# R	# T	# R	# T
Single Robot	Single Tag	1	1	N_A	1	N_A	1	1
	Multi Tag	1	N_T	N_A	1	N_A	1	N_T
Multi Robot	Single Tag	N_R	1	0	1	1	N_R-1	1
	Multi Tag	N_R	N_T	0	1	N_T	N_R-1	N_T

B. Multi-robot Multi-tag Positioning

In this setting, the objective is to predict the relative poses of all neighbour robots $\{\hat{\mathbf{P}}_{B_i}^{B_r} = \begin{bmatrix}\hat{\mathbf{R}}_{B_i}^{B_r} & \hat{\mathbf{t}}_{B_i}^{B_r} \\ 0 & \mathbf{1}\end{bmatrix} | i, r \in \mathcal{S}_\mathcal{R}, i \neq r\}$ with respect to a reference robot R_r , which can be any robot in the group. There are three primary types of constraints: extrinsic constraints, distance constraints, and bearing constraints, which are described as:

$$\begin{aligned} \mathbf{t}_{T_{i,m}}^{B_r} &= \hat{\mathbf{R}}_{B_i}^{B_r}\mathbf{t}_{T_{i,m}}^{B_i} + \hat{\mathbf{t}}_{B_i}^{B_r} + \boldsymbol{\eta}_e \\ d_{j,n}^{i,m} &= \| \mathbf{t}_{T_{i,m}}^{B_r} - \mathbf{t}_{T_{j,n}}^{B_r} \| + \eta_d \\ \mathbf{b}_{B_i}^{B_r} &= \frac{\hat{\mathbf{t}}_{B_i}^{B_r}}{\| \hat{\mathbf{t}}_{B_i}^{B_r} \|} + \boldsymbol{\eta}_b \end{aligned} \tag{1}$$

The relative position of the m-th tag attached to R_i is denoted as $\mathbf{t}_{T_{i,m}}^{B_r}$, which is calculated by the tag extrinsics $\mathbf{t}_{T_{i,m}}^{B_i}$ and the predicted pose. The $d_{j,n}^{i,m}$ refers to the pairwise distance between the m-th tag on the i-th robot and the n-th tag on the j-th robot. The $\mathbf{b}_{B_i}^{B_r}$ denotes the bearing vector of robot R_i observed by robot R_r in its local frame. The inter-robot bearing vectors can be produced through either rule-based hard matching or learning-based soft assignment. We evaluate the performance differences between distinct matching strategies in the experimental analysis in Appendix F. The $\boldsymbol{\eta}_e$, η_d, and $\boldsymbol{\eta}_b$ correspond to the bias and noise terms associated with the extrinsic, distance, and bearing measurements, respectively.

C. Unified Formulation

The single-tag and non-bearing settings still fall within the general category of multi-tag with bearing formulation we

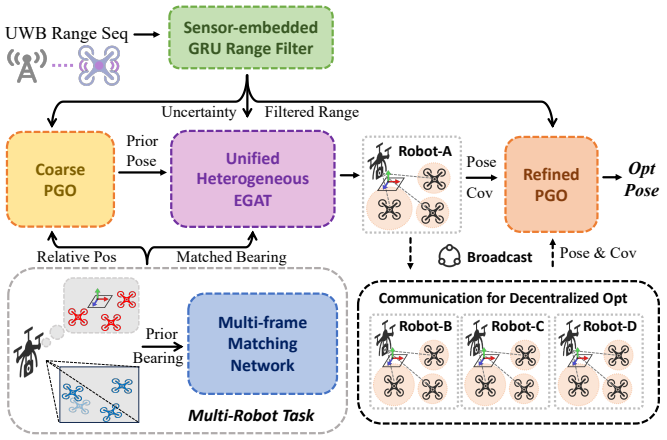


Fig. 3: The framework of our proposed **AnyAmber**, unified anonymous bearing and range based position tracking system.

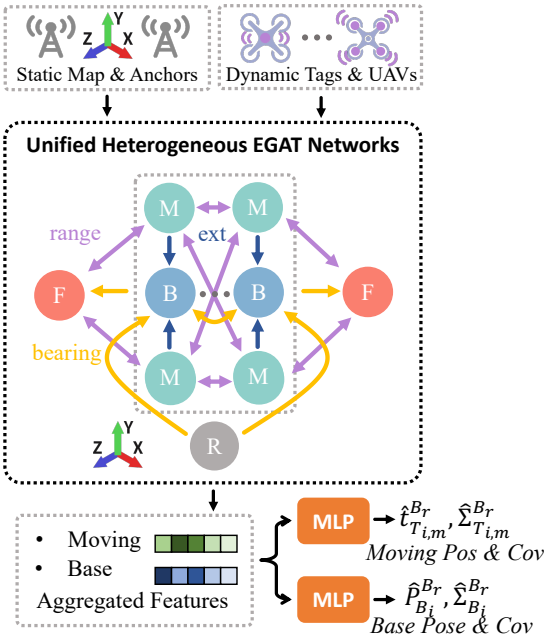


Fig. 4: The heterogeneous unified EGAT localization network, incorporating four node types (Ref, Base, Fixed, Moving) and three distinct edge types (ext, range, bearing). The GNN-aggregated moving and base features are utilized to estimate the corresponding pose and covariance.

discussed before, differing only by the degeneration of their respective constraints. Thus, forming a unified formulation requires addressing two key issues: (1) converting single-robot global localization in the map frame into multi-robot relative localization, and (2) incorporating the information from static anchors fixed in the map frame.

To overcome these issues, we reinterpret the map frame in single-robot localization as a virtual reference robot. The anchors defined in the map frame are considered as tags mounted on the reference robot, whose map-frame locations serve as the extrinsic parameters of the UWB tags. Through this reformulation, the single-robot positioning task is converted into a multi-robot relative positioning problem involving a reference robot (with multiple tags) and a target robot to be estimated. Tab. I summarizes the configurations of the four representative localization tasks, including robots, anchors, and tags in each scenario before and after the reformulation.

IV. Network Architecture

Our system overview is presented in Fig. 3, which consists of a GRU range filter network, a multi-frame matching module, the key unified heterogeneous EGAT network and two differentiable PGO blocks.

Given a sequence of UWB ranging inputs and the corresponding sensor feature embeddings, the Sensor-Embedded GRU performs filtering and ranging uncertainty estimation, and the resulting features are fed into the downstream Coarse PGO (CPGO), EGAT, and Refined PGO (RPGO) modules. In multi-robot scenarios with visual detections, the matched bearing information and relative positions are extracted from anonymous visual observations by a GNN-based multi-frame matching network. With the comprehensive inputs of filtered ranges, matched bearings and pose priors, the core EGAT estimates both the poses and covariances of the robots. The CPGO and RPGO modules perform prior pose correction and output pose refinement at different stages, respectively.

Our system operates in a decentralized manner when applied to multi-robot settings, where robots communicate only by broadcasting predicted poses and associated covariances after EGAT, and subsequently perform RPGO independently, leading to minimal communication overhead. In the following sections, we provide a detailed introduction to the design and functionality of each submodule.

A. Heterogeneous Unified EGAT Localization Network

Building upon the unified multi-robot multi-tag formulation introduced in Section III-C, we propose a heterogeneous graph neural network with four node types and three edge types, with details provided in Appendix A. Different from the approach in [13], which employs a standard GAT network with all information stored in node features, we follow the approach in [34] and adopt an EGAT network. In this design, the estimated quantities are represented as nodes, while geometric constraints such as range and bearing are encoded as edges, providing a more structured and interpretable problem formulation. The network architecture is illustrated in Fig. 4.

1) Graph Nodes: As outlined before, we unify the map frame in the single-robot scenario with the reference robot in the multi-robot scenario, which is denoted here as *Ref*. Here the *Ref* feature simply represents the identity pose $\mathbf{P}_{B_r}^{B_r}$ expressed in its own reference frame. Similarly, the static anchors are unified with the tags on the reference robot, denoted as *Fixed*, with the feature $\mathbf{t}_{T_{r,n}}^{B_r}$ representing the tag positions. The *Base* node denotes the pose $\hat{\mathbf{P}}_{B_i}^{B_r}$ of the robot being estimated, while the *Moving* node corresponds to the position $\hat{\mathbf{t}}_{T_{i,m}}^{B_r}$ of the tag attached to it. It is noteworthy that *Ref* and *Base* nodes, as well as *Fixed* and *Moving* nodes, could theoretically be further merged into a more unified

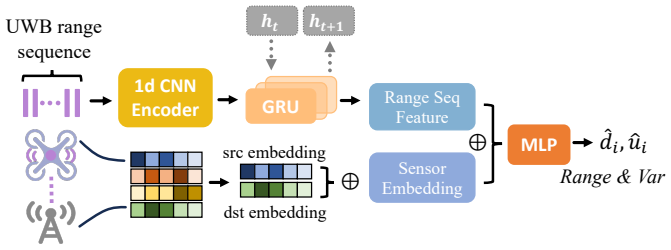


Fig. 5: The Sensor-Embedded GRU range filter network.

architecture. We retain their distinction because reference nodes and estimated nodes may vary in confidence and feature aggregation behaviors. Representing them as separate node types yields superior performance.

2) Graph Edges: The unified localization formulation incorporates three constraint types: extrinsic, bearing, and range constraints, which motivate the three corresponding edge types in the GNN. In bearing observations, we typically establish a bearing edge from the reference robot to the target robot, namely *Ref*-to-*Base*. When a bearing observation exists between neighboring robots, a *Base*-to-*Base* edge can be derived. Similarly, a *Base*-to-*Fixed* edge is derived when a robot observes static anchors. For range measurements, *range* edges are subdivided into *Moving*-to-*Fixed*, *Fixed*-to-*Moving*, and *Moving*-to-*Moving* types based on whether the range node corresponds to an estimated node, and the weights of these range edges are mutually independent.

3) Feature Aggregation: Following the same message aggregation mechanism in [34], we finally extract the node features of *Moving* and *Base* to predict their poses and associated uncertainties, with details presented in Appendix A.

B. Sensor-Embedded GRU Network

It is widely recognized that UWB ranging can be impaired by noise and systematic bias, often caused by antenna deployment and inter-node signal interference, giving rise to a non-zero-mean, non-Gaussian ranging pattern [5, 35]. More critically, the noise and bias are typically time-varying and node-pair-dependent, which implies that a generic filter with fixed parameters often yields suboptimal denoising performance.

To tackle this issue, a sensor-embedding–augmented GRU network is introduced to facilitate tailored and adaptive calibration, consisting of a standard GRU module to capture the general temporal ranging pattern for effective suppression of ranging noise during pretraining, and a sensor embedding table to model the scenario-specific pair-wise sensor characteristics for sensor-coupled ranging biases correction during finetuning. A batch of UWB ranging measurements within a given window is first processed by a 1D CNN to extract temporal ranging features. Subsequently, a three-layer GRU network integrates the historical hidden representations with the current ranging features to generate more expressive range embeddings. To model the distinct characteristics of each sensor pair, the unique IDs of the transmitting and receiving sensors are used to fetch their node embeddings from a pre-trained embedding table. The embeddings from sensor pairs and range sequences are then added together to generate the pairwise sensor embeddings, followed by the MLP to predict the filtered range and its corresponding variance.

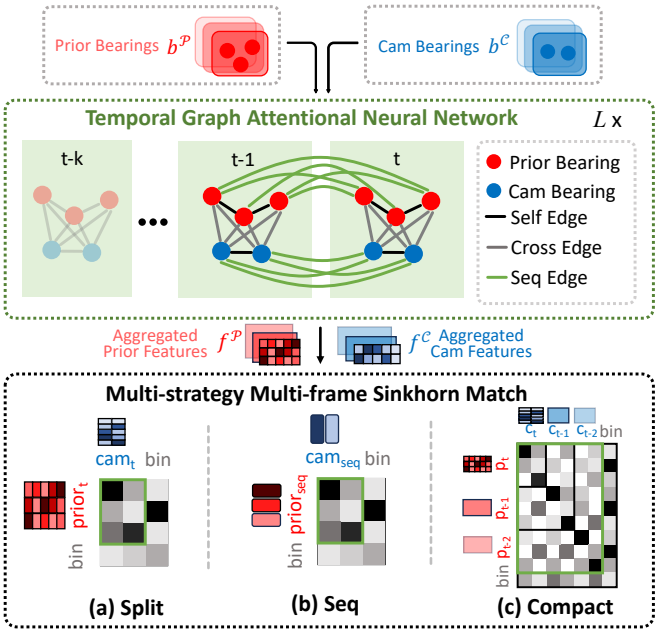


Fig. 6: Multi-frame temporal matching network.

C. Multi-frame Matching Network

Motivated by the research about learnable multi-robot matching [13], we solve the data association problem between ID-aware priors and anonymous visual detections using graph-based approach. To effectively capture temporal dependencies, we extend the single-frame matching paradigm to a multi-frame setting by introducing temporal GNN-based networks (encoder module) and multi-frame Sinkhorn partial assignments (decoder and matching module). The architecture of our graph match net is illustrated in Fig. 6.

1) Temporal Attentional Graph Neural Network: The initial node features of priors $\mathcal{P}$ and camera detections $\mathcal{C}$ are represented by their high-dimensional embeddings ${}^{(0)}\mathbf{f}_i$ projected from raw bearings via a shared MLP. We design three distinct types of cross-attention edges in the GNN: self-edges, cross-edges, and sequential-edges, enabling a more profound integration of the multi-robot formation characteristics (self-edge), the similarity cues from candidate bearing measurements (cross-edge), and the temporal context (seq-edge). The lifted node embeddings undergo three successive aggregation stages: (1) self-edge aggregation with nodes from the same set at the same time step, (2) cross-edge aggregation with nodes from the other set at the same time step, and (3) sequential aggregation with nodes from different time steps within the same set. Note that residual connections are utilized at both intra-layer and inter-layer levels to facilitate feature propagation.

$$
\begin{aligned}
{}^{(l)}\mathbf{f}_i^{self} &= {}^{(l)}\mathbf{f}_i + {}^{(l)}f_{self}([{}^{(l)}\mathbf{f}_i \,||\, {}^{(l)}\mathbf{m}_{\varepsilon_{self}}]) \\
{}^{l}\mathbf{f}_i^{cross} &= {}^{(l)}\mathbf{f}_i^{self} + {}^{(l)}f_{cross}([{}^{(l)}\mathbf{f}_i^{self} \,||\, {}^{(l)}\mathbf{m}_{\varepsilon_{cross}}]) \\
{}^{(l+1)}\mathbf{f}_i &= {}^{(l)}\mathbf{f}_i^{cross} + {}^{(l)}f_{seq}([{}^{(l)}\mathbf{f}_i^{cross} \,||\, {}^{(l)}\mathbf{m}_{\varepsilon_{seq}}])
\end{aligned} \tag{2}
$$

where $||$ denotes concatenation. ${}^{(l)}\mathbf{f}_i$ is the bearing embedding of layer l. The message aggregation along self-edge ε_{self}, cross-edge ε_{cross} and seq-edge ε_{seq} are represented by ${}^{(l)}\mathbf{m}_{\varepsilon_{self}}$, ${}^{(l)}\mathbf{m}_{\varepsilon_{cross}}$ and ${}^{(l)}\mathbf{m}_{\varepsilon_{seq}}$ respectively. ${}^{(l)}f_{self}$, ${}^{(l)}f_{cross}$ and ${}^{(l)}f_{seq}$ are MLPs to merge attentional features and own embeddings, ensuring that the feature dimensions remain consistent before and after the aggregation.

After extensive message passing through multiple layers of the GNN, we obtain two bearing feature sets $\mathbf{f}^{\mathcal{P}}$ and $\mathbf{f}^{\mathcal{C}}$, with enhanced spatiotemporal representation capabilities.

2) Multi-frame Partial Assignment: We introduce three distinct strategies for performing multi-frame partial assignment, namely *Split*, *Sequential* and *Compact*. The *Split* method divides the multi-frame bearing set into individual frames and performs frame-to-frame matching between the prior set and the camera set. The *Sequential* method formulates multi-frame matching as a two-stage sequential matching problem. The *Compact* method constructs a single score matrix ${}^{k}\mathbf{S} \in \mathbb{R}^{(N_R * N_W)\times(N_C * N_W)}$ by stacking bearing features from all frames, enabling joint iterative optimization over the entire sequence in a single step. More details of these three assignment algorithms can be found in Appendix B.

D. Hierarchical Differentiable PGO

To enhance the overall position tracking precision, we adopt a hierarchical differentiable PGO architecture composed of CPGO and RPGO. These two components share identical constraint types, but operate on different input data. There are four major constraints involved in PGO: pose prior constraint, extrinsic constraint, UWB ranging constraint and mutual state constraint. The CPGO module integrates the previous frame's predictions with the current frame observations to provide a refined pose prior for the EGAT localization network. Subsequently, the RPGO module takes the pose predictions and uncertainty estimates generated by the EGAT network to optimize the final pose output.

In our implementation, we leverage Theseus [36] for differentiable nonlinear optimization. The gradients backpropagated from this optimization backend facilitate the EGAT network's learning of predictive uncertainties.

V. Training Settings

A. Loss Functions

To jointly improve the performance of matching, filtering, and localization, we define a composite loss function to supervise the network training:

$$
\mathcal{L} = \mathcal{L}_{Match} + \lambda_p \mathcal{L}_{Pose} + \lambda_r \mathcal{L}_{Range} \tag{3}
$$

where λ_r and λ_p are weighting factors for different loss terms.

1) Match Loss: Following N_I Sinkhorn iterations, we obtain the score matrix ${}_{(N_I)}\bar{\mathbf{S}}$, which may represent frame-to-frame, sequence-to-sequence or compact matching. Suppose the matrix has $N+1$ rows and $M+1$ columns. We supervise the specific item in the iterated score matrix:

$$
\mathcal{L}_{Match} = -\sum_{(i,j)\in\pi} {}_{(N_I)}\bar{\mathbf{S}}_{i,j} - \sum_{i\in\mu} {}_{(N_I)}\bar{\mathbf{S}}_{i,M} - \sum_{j\in\upsilon} {}_{(N_I)}\bar{\mathbf{S}}_{N,j} \tag{4}
$$

where π is the set of matching bearings, μ and υ denote the set of unmatched candidates, derived from the ground truth labels. The first term encourages higher scores for correct correspondences, while the latter two terms promote assigning incorrect matches to the dustbin row and column.

2) Pose Loss: Pose supervision is achieved through two losses: a maximum-likelihood (ML) loss on the EGAT-predicted poses and variances, and a Mean Squared Error (MSE) loss on the RPGO-refined poses, whose gradients are further propagated backward to facilitate end-to-end training of the upstream network.

$$
\begin{aligned}
\mathcal{L}_{Pose} &= {}^{ML}\mathcal{L}_{Pose} + \lambda_m \cdot {}^{MSE}\mathcal{L}_{Pose} \\
{}^{ML}\mathcal{L}_{Pose} &= \frac{1}{N_S}\sum_{i\in\mathcal{S}} (\lambda_d log(det(\hat{\mathbf{\Sigma}}_i)) \\
&\quad + (\mathbf{e}_i - \hat{\mathbf{e}}_i)^T \hat{\mathbf{\Sigma}}_i^{-1} (\mathbf{e}_i - \hat{\mathbf{e}}_i)) \\
{}^{MSE}\mathcal{L}_{Pose} &= \frac{1}{N_S}\sum_{i\in\mathcal{S}} (\| \mathbf{t}_i - \hat{\mathbf{t}}_i \|^2 + \lambda_q \| \mathbf{q}_i - \hat{\mathbf{q}}_i \|^2)
\end{aligned} \tag{5}
$$

where $\mathcal{S}$ is the set of robots to be estimated and N_S is the number of variables in this set. $\hat{\mathbf{e}}_i \in \mathbb{R}^3$ represents the predicted state of the i-th robot, specifically its estimated position or Euler angles (converted from quaternions). While $\mathbf{e}_i$ and $\hat{\mathbf{\Sigma}}_i \in \mathbb{R}^{3\times3}$ are its corresponding ground truth and predicted covariance matrix. For the MSE loss term, the $\hat{\mathbf{t}}_i$ and $\hat{\mathbf{q}}_i$ refer to the optimized position and quaternion from the RPGO module. λ_d, λ_m and λ_q are the coefficients that balance the loss weights of the determinant term, the MSE term and the quaternion term.

3) Range Loss: To enable the GRU network to achieve improved range filtering performance and uncertainty estimation, we adopt the ML loss, but reduce it to a one-dimensional form.

$$
\mathcal{L}_{Range} = \frac{1}{N_D}\sum_{i\in\mathcal{D}} (\lambda_d log(\hat{u}_i) + \frac{(d_i - \hat{d}_i)^2}{\hat{u}_i}) \tag{6}
$$

where $\mathcal{D}$ is the set of distance measurements to be filtered and N_D is the number of ranging in this set. $\hat{d}_i$, $\hat{u}_i$ and d_i are the filtered distance, uncertainty prediction and ground truth distance, respectively.

B. Joint Pretraining and Scene-specific Finetuning

During pretraining, for efficiency considerations, the CPGO and RPGO modules are omitted. The EGAT localization network, the sensor-agnostic GRU network, and the GNN matching network are pretrained independently, forming transferable general knowledge for localization, filtering, and matching.

TABLE II: Multi-scene multi-task positioning RMSE (m). All test scenarios are strictly unseen during pre-training. Test scenarios marked with * indicate that both the scenario itself and its scenario type have never been encountered during pre-training. The simulation scenarios are denoted by the *sim-* prefix. The best positioning result is highlighted in bold, the second-best is marked with underlining, the - indicates that the method is not applicable to the corresponding localization task.

Method		Learned	Unified	Single-Robot				Multi-Robot			
				Single-Tag		Multi-Tag		Single-Tag		Multi-Tag	
				sim-circle	square*[7]	tunnel drive	tunnel fly	lab-los*[14] (w. bearing)	lab-nlos*[14] (w. bearing)	sim-spiral (w. bearing)	plane*[32] (wo. bearing)
Individual Training	VIUNet[7] (SS)	✓	✗	1.08	0.12	-	-	-	-	-	-
	NRIO[13] (SM)	✓	✗	-	-	0.18	0.26	-	-	-	-
	Mr.Virgil[14] (MS)	✓	✗	-	-	-	-	0.63	**0.14**	-	-
Optimization -based	MURP[32] (MM)	✗	✗	-	-	-	-	-	-	-	1.44
	PGO	✗	✓	0.63	0.14	0.59	0.44	0.77	0.17	5.66	1.78
Joint Training	**AnyAmber**	✓	✓	**0.17**	**0.04**	**0.15**	**0.11**	**0.09**	0.15	**0.11**	**0.36**

Although various localization tasks are unified under a multi-robot multi-tag formulation, some degrees of freedom are still underdetermined in specific settings, such as the ambiguity of orientation in single-tag scenarios without bearing information. To facilitate joint pretraining, we adopt a unified output representation and apply a state mask to supervise the estimated states selectively.

Different localization tasks often involve diverse scene-dependent factors, such as varying UWB node distributions, heterogeneous sensor configurations, and differing noise characteristics. As a result, it is necessary to perform fine-tuning tailored to the characteristics of each individual scenario.

In the fine-tuning stage, we augment the GRU network with sensor embeddings to learn scene-specific ranging noise properties. Furthermore, the CPGO and RPGO modules are integrated into the end-to-end training framework, allowing gradients from the optimization procedure to backpropagate and supervise the EGAT network in covariance learning.

The model has encapsulated generalized localization knowledge after pretraining, enabling effective fine-tuning for specific scenes with limited data. In practice, we observe that fine-tuning on merely one trajectory yields substantial performance improvements in the target scenario.

VI. Experiment

A. Experimental Settings and Datasets

We developed a comprehensive multi-scenario localization dataset comprising both simulation and real-world experiments. The dataset spans diverse conditions, including varying numbers of robots, anchors, and UWB tags, alongside diverse robot trajectories and environmental influences. It contains 113 distinct scenes, 317 trajectories, and 210589 sampled frames, among which 108 scenes and 225 trajectories are used for joint pretraining, encompassing configurations with 1 to 16 robots, 0 to 36 anchors, and 1 to 6 tags. The details of our unified dataset and settings are presented in Appendix C.

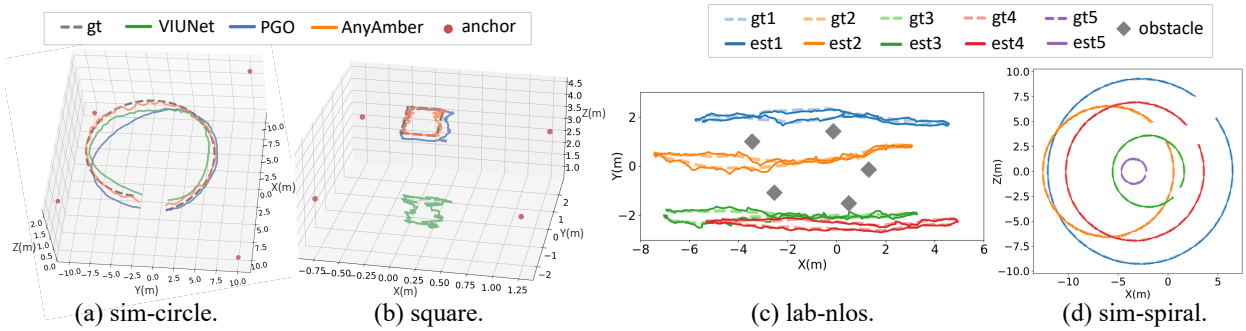
We evaluate the RMSE localization errors of our method against general and task-specific approaches. Detailed descriptions of the baselines are provided in the Appendix D. Since the learning-based baselines are designed with specialized network architectures, we train each of them individually using as much training data as possible on its corresponding task. Our model, by contrast, is first pre-trained on the unified dataset and then fine-tuned in each specific test scenario using only a single trajectory of approximately 150 frames.

B. Multi-scene Position Tracking Accuracy

The comprehensive experimental results are presented in Tab. II. Representative experimental trajectories are visualized in Fig. 7, with additional results available in Appendix E. Our method not only adapts to various positioning tasks, including single-robot, multi-robot, single-tag, and multi-tag configurations, but also achieves optimal or near-optimal performance in each specific scenario.

In the single-robot scenario, the error metrics are calculated in the nominal frame. Fig. 7 (a)(b) compares the trajectories of ground-truth, our method's prediction, and baseline predictions. Our model outperforms all learning-based and optimization-based approaches owing to the strong foundation established by the unified and learnable architecture.

For the multi-robot case in Fig. 7 (c)(d), we visualize only our predictions against the ground truth for clarity, where neighbor robot trajectories are plotted in the world frame by transforming their estimated relative poses using the world pose of the reference robot. The error metrics in Tab. II are calculated in the local frame of the reference robot. Under the *lab-los* scenario where the robots move at relatively high speeds, Mr. Virgil's accuracy deteriorates significantly once pseudo odometry is removed, due to its heavy reliance on prior estimates. In contrast, our method limits the localization error to under 10 cm. In the multi-robot multi-tag scenario *plane*, the z-axis is weakly constrained and no bearing measurements are



(a) sim-circle. (b) square. (c) lab-nlos. (d) sim-spiral.

Fig. 7: Visualization of representative trajectory estimates in single-robot (a–b) and multi-robot (c–d) scenarios.

TABLE III: Ablation study on system architecture evaluated by localization RMSE (m)

EGAT	GRU	Coarse PGO	Refined PGO	Single-Robot				Multi-Robot				Avg.
				Single-Tag		Multi-Tag		Single-Tag		Multi-Tag		
				sim-circle	square*	tunnel drive	tunnel fly	lab-los* (w. bearing)	lab-nlos* (w. bearing)	sim-spiral (w. bearing)	plane* (wo. bearing)	
✓				0.44	**0.04**	0.45	0.24	0.18	0.20	0.27	3.13	0.62
✓		✓	✓	0.28	**0.04**	**0.15**	**0.11**	0.12	0.17	0.13	1.23	0.28
✓	✓		✓	**0.16**	**0.04**	0.16	**0.11**	**0.09**	0.16	0.13	0.82	0.21
✓	✓	✓		0.35	**0.04**	0.21	0.13	0.37	0.27	0.20	0.45	0.25
✓	✓	✓	✓	0.17	**0.04**	**0.15**	**0.11**	**0.09**	**0.15**	**0.11**	**0.36**	**0.15**

available, making the task considerably challenging. Despite this, our method maintains a relatively low localization error, with errors reduced to 25% or less of those of the baselines, demonstrating strong robustness.

C. Ablation Study

We conduct ablation studies on different network architectural configurations to demonstrate the functionality of each submodule, with quantitative results reported in Tab. III. The Sensor-Embedded GRU can model non-Gaussian UWB ranging characteristics and correct sensor biases effectively, particularly in scenarios with large ranging noise and bias, reducing the mean localization error by over 40%. In most scenarios, systems without Coarse PGO can still achieve comparable localization accuracy. However, in more challenging scenarios, such as *plane*, the correction of prior poses via Coarse PGO provides more accurate initialization, thereby promoting network convergence. Refined PGO functions as the backend of the system, combining sensor observations with EGAT-derived pose and uncertainty estimates, which substantially enhances localization performance for most localization tasks.

D. Network Generalization Ability Analysis

In this section, we investigate the generalization capability of the model through experiments under different numbers of training scenes and varying robot counts.

1) Training Trajectory Numbers: We evaluate the localization performance of pretrained and non-pretrained models in a single-robot, single-tag simulation scenario, fine-tuning each model with different numbers of trajectories. As illustrated in Fig. 2, the pretrained model, after being fine-tuned on merely one trajectory, generalizes effectively to unseen scenarios and yields significantly lower localization error than the non-pretrained model even when the latter is fine-tuned on five trajectories. This highlights that the joint pretraining process endows the model with transferable multi-scene localization knowledge, substantially alleviating data demands during downstream deployment.

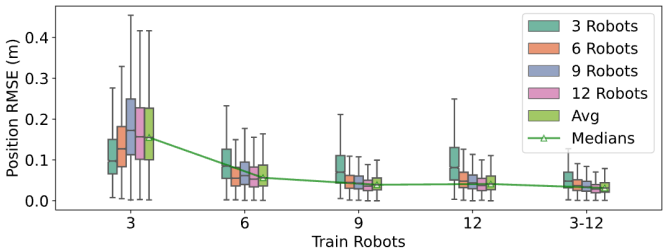


Fig. 8: RMSE box plot of varying robot numbers experiment.

2) Robot Numbers: To evaluate the model's ability to handle different robot scales, we train models from scratch (without pretraining) on datasets containing 3, 6, 9, 12, and 3–12 robots, and test each model across scenes with varying numbers of robots. According to the error box plots in Fig. 8, the model trained on the mixed 3–12 robot dataset exhibits strong generalization across test scenarios with varying robot numbers, and can even surpass the models trained specifically for a fixed robot count on its corresponding scenario.

TABLE IV: Computational time (ms) of each submodule and the full pipeline under varying robot scales.

#Robot	MatchNet	GRU	CPGO	EGAT	RPGO	**Total**
4	21.2	0.9	0.2	11.8	0.2	34.3
8	20.7	0.9	0.6	11.5	0.5	34.3
12	21.2	0.9	1.7	11.5	1.6	36.9
16	22.0	0.9	5.0	11.6	3.7	43.2

E. Computational Efficiency and Scalability

To ensure real-time inference, we implement the computationally heavy CPGO and RPGO modules in C++ using Ceres Solver, and leverage pybind11 for high-performance calls from Python during inference. The neural network inference is accelerated via GPU, batch processing, and parallelization. As computation time is summarized in Tab. IV, the system achieves an output rate of at least 20 Hz, even when scaling to 16 robots. We also vary the observation neighborhood in a 16-robot setup; the matching network runtime remains nearly unchanged (20.94 ms for size 6 vs. 20.97 ms for size 15), highlighting the model's scalability.

VII. Conclusion

In this work, we propose a unified modeling formulation and network architecture capable of handling a wide range of anonymous bearing and ranging based localization tasks, including single-robot, multi-robot, single-tag, and multi-tag settings. We construct a comprehensive multi-task dataset for pretraining, integrating simulation and real-world data, self-collected and public datasets. The pretrained model can be adapted to unseen scenarios with only a single fine-tuning trajectory and achieves superior or comparable localization accuracy to task-specific architectures.

Nevertheless, due to the output-level fusion of the matching and localization networks in current methods, positional priors remain indispensable, preventing our approach from scaling toward global localization in a fully prior-free manner. Notably, our architecture readily supports extension to a truly integrated end-to-end design (e.g., feeding GRU/MatchNet latent features directly into EGAT). As part of our future efforts, we will explore prior-free global localization by leveraging odometry, multi-frame observations, and feature-level end-to-end fusion.

Acknowledgments

This work was supported by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant No. JYB2025XDXM208, the National Nature Science Foundation of China under Grant No. 62522317 and 62503144, the Joint Funds of the National Natural Science Foundation of China under Grant No. U24A20128, Zhejiang Provincial Natural Science Foundation of China under Grant No. LD25F030001, the Key R&D Program of Hangzhou under Grant No. 2024SZD1A30.

References

[1] Aldo Jaimes, Srinath Kota, and Jose Gomez. An approach to surveillance an area using swarm of fixed wing and quad-rotor unmanned aerial vehicles uav (s). In *2008 IEEE International Conference on System of Systems Engineering*, pages 1–6. IEEE, 2008.

[2] Yang Qi, Yisheng Zhong, and Zongying Shi. Cooperative 3-d relative localization for uav swarm by fusing uwb with imu and gps. In *Journal of Physics: Conference Series*, volume 1642, page 012028. IOP Publishing, 2020.

[3] SungTae Moon, YeonJu Choi, DoYoon Kim, Myeonghun Seung, and HyeonCheol Gong. Outdoor swarm flight system based on rtk-gps. *Journal of KIISE*, 43(12):1315–1324, 2016.

[4] James A Preiss, Wolfgang Honig, Gaurav S Sukhatme, and Nora Ayanian. Crazyswarm: A large nano-quadcopter swarm. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3299–3304. IEEE, 2017.

[5] Haodong Jiang, Wentao Wang, Yuan Shen, Xinghan Li, Xiaoqiang Ren, Biqiang Mu, and Junfeng Wu. Efficient planar pose estimation via uwb measurements. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1954–1960. IEEE, 2023.

[6] Bo Yang, Jun Li, Zhanpeng Shao, and Hong Zhang. Robust uwb indoor localization for nlos scenes via learning spatial-temporal features. *IEEE Sensors Journal*, 22(8):7990–8000, 2022.

[7] Peng-Yuan Kao, Hsiu-Jui Chang, Kuan-Wei Tseng, Timothy Chen, He-Lin Luo, and Yi-Ping Hung. Viunet: Deep visual–inertial–uwb fusion for indoor uav localization. *IEEE Access*, 11:61525–61534, 2023.

[8] Zahra Arjmandi, Jungwon Kang, Gunho Sohn, Costas Armenakis, and Mozhdeh Shahbazi. Deepcovpg: Deep-learning-based dynamic covariance prediction in pose graphs for ultra-wideband-aided uav positioning. In *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*, pages 3836–3843. IEEE, 2024.

[9] HongZhou Zheng, Fei Meng, and JinRu Han. Enhanced ultra wide band indoor localization in non line of sight environments using graph attention networks. In *2023 7th CAA International Conference on Vehicular Control and Intelligence (CVCI)*, pages 1–6. IEEE, 2023.

[10] Haoran Huang, Chuhao Chen, Kailong Wang, Ruini Wang, and Shizhao Zhao. A novel gnn-based method with light weight attention augmentation for indoor positioning. In *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, pages 138–141. IEEE, 2024.

[11] Sizhen He, Bo Yang, Tao Liu, and Hong Zhang. Multi-tag uwb localization with spatial-temporal attention graph neural network. *IEEE Transactions on Instrumentation and Measurement*, 2024.

[12] Sizhen He, Bo Yang, Tao Liu, and Jun Li. Graph network-based uwb localization via learning spatial-temporal and geometric features. *IEEE Communications Letters*, 2025.

[13] Si Wang, Bingqi Shen, Fei Wang, Yanjun Cao, Rong Xiong, and Yue Wang. Neural ranging inertial odometry. In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9377–9383. IEEE, 2025.

[14] Si Wang, Zhehan Li, Jiadong Lu, Rong Xiong, Yanjun Cao, and Yue Wang. Mr. virgil: Learning multi-robot visual-range relative localization. In *2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 15414–15421. IEEE, 2025.

[15] Hanying Zhao, Lingwei Xu, Yi Li, Feiyang Wen, Haoran Gao, Changwu Liu, Jincheng Yu, Yu Wang, and Yuan Shen. Robust and scalable multi-robot localization using stereo uwb arrays. *IEEE Transactions on Robotics*, 2025.

[16] Chenyang Liang, Liangming Chen, Baoyi Cui, and Jie Mei. 3-d relative localization for multi-robot systems with angle and self-displacement measurements. *The International Journal of Robotics Research*, 2025.

[17] Zhiren Xun, Jian Huang, Zhehan Li, Zhenjun Ying, Yingjian Wang, Chao Xu, Fei Gao, and Yanjun Cao. Crepes: Cooperative relative pose estimation system. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5274–5281. IEEE, 2023.

[18] Edwin Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE international conference on robotics and automation*, pages 3400–3407. IEEE, 2011.

[19] Xudong Yan, Heng Deng, and Quan Quan. Active infrared coded target design and pose estimation for multiple objects. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6885–6890. IEEE, 2019.

[20] Hao Xu, Luqi Wang, Yichen Zhang, Kejie Qiu, and Shaojie Shen. Decentralized visual-inertial-uwb fusion for relative state estimation of aerial swarm. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 8776–8782. IEEE, 2020.

[21] Hao Xu, Yichen Zhang, Boyu Zhou, Luqi Wang, Xinjie Yao, Guotao Meng, and Shaojie Shen. Omni-swarm: A decentralized omnidirectional visual–inertial–uwb state estimation system for aerial swarms. *Ieee transactions on robotics*, 38(6):3374–3394, 2022.

[22] Hao Xu, Peize Liu, Xinyi Chen, and Shaojie Shen. D^2 SLAM: Decentralized and distributed collaborative visual-inertial slam system for aerial swarm. *IEEE Transactions on Robotics*, 2024.

[23] Bowen Cheng, Alex Schwing, and Alexander Kirillov. Per-pixel classification is not all you need for semantic segmentation. *Advances in neural information processing systems*, 34:17864–17875, 2021.

[24] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar. Masked-attention mask transformer for universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1290–1299, 2022.

[25] Jitesh Jain, Jiachen Li, Mang Tik Chiu, Ali Hassani, Nikita Orlov, and Humphrey Shi. Oneformer: One transformer to rule universal image segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2989–2998, 2023.

[26] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 5294–5306, 2025.

[27] Jiaxin Li, Yingcai Bi, Kun Li, Kangli Wang, Feng Lin, and Ben M Chen. Accurate 3d localization for mav swarms by uwb and imu fusion. In *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, pages 100–105. IEEE, 2018.

[28] Yanjun Cao, Chenhao Yang, Rui Li, Alois Knoll, and Giovanni Beltrame. Accurate position tracking with a single uwb anchor. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 2344–2350. IEEE, 2020.

[29] Hashim A Hashim, Abdelrahman EE Eltoukhy, Kyriakos G Vamvoudakis, and Mohammed I Abouheaf. Nonlinear deterministic observer for inertial navigation using ultra-wideband and imu sensor fusion. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3085–3090. IEEE, 2023.

[30] Kailai Li, Ziyu Cao, and Uwe D Hanebeck. Continuous-time ultra-wideband-inertial fusion. *IEEE Robotics and Automation Letters*, 8(7):4338–4345, 2023.

[31] Kailai Li. On embedding b-splines in recursive state estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 2025.

[32] Andrew Fishberg, Brian Quiter, and Jonathan P How. Murp: Multi-agent ultra-wideband relative pose estimation with constrained communications in 3d environments. *IEEE Robotics and Automation Letters*, 2024.

[33] Yijun Yuan, Zhuoguang Chen, Kenan Li, Weibang Wang, and Hang Zhao. Slam-former: Putting slam into one transformer. *arXiv preprint arXiv:2509.16909*, 2025.

[34] Thomas Monninger, Julian Schmidt, Jan Rupprecht, David Raba, Julian Jordan, Daniel Frank, Steffen Staab, and Klaus Dietmayer. Scene: Reasoning about traffic scenes using heterogeneous graph neural networks. *IEEE Robotics and Automation Letters*, 8(3):1531–1538, 2023.

[35] Thien-Minh Nguyen, Abdul Hanif Zaini, Chen Wang, Kexin Guo, and Lihua Xie. Robust target-relative localization with ultra-wideband ranging and communication. In *2018 IEEE international conference on robotics and automation (ICRA)*, pages 2312–2319. IEEE, 2018.

[36] Luis Pineda, Taosha Fan, Maurizio Monge, Shobha Venkataraman, Paloma Sodhi, Ricky TQ Chen, Joseph Ortiz, Daniel DeTone, Austin Wang, Stuart Anderson, et al. Theseus: A library for differentiable nonlinear op-

timization. *Advances in Neural Information Processing Systems*, 35:3801–3818, 2022.

[37] Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li, and Zheng Zhang. Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315*, 2019.

[38] Jun Hyeok Choe and Inwook Shim. Addressing relative pose impact on uwb localization: Dataset introduction and analysis. In *2024 24th International Conference on Control, Automation and Systems (ICCAS)*, pages 305–308. IEEE, 2024.