Disturbance-Aware Underwater Visual-Inertial Odometry via Learned Dynamics and External Force Estimation

Yazan Maalla[1], Zein Alabedeen Barhoum[1], Maxim Popov[1] and Sergey Kolyubin[1]

Abstract— **Reliable state estimation for autonomous underwater robots is challenging due to GPS denial, poor visibility, and complex hydrodynamic disturbances. Conventional visual-inertial odometry (VIO) ignores actuation dynamics, misinterpreting external forces as sensor noise and causing drift.**

We present a dynamics-informed visual-inertial odometry framework that integrates physically consistent learned dynamics with visual-inertial fusion for improved underwater state estimation. A Port-Hamiltonian Neural ODE learns energy-conserving dynamics directly from data. This learned model is preintegrated into a tightly coupled factor graph alongside visual and inertial constraints, enabling explicit separation of commanded motion from disturbance-induced effects. A dedicated dynamics residual attributes motion discrepancies to an external force variable, jointly optimized with the trajectory to provide interpretable force estimates without corrupting pose estimation.

Evaluation on real-world underwater datasets demonstrates consistent accuracy improvements over baseline VIO with less than 4% computational overhead. In controlled simulation experiments with known disturbances, the framework reduces trajectory error by over 60% and reconstructs applied forces with near-unity correlation (R=0.99). To our knowledge, this is the first integration of physically consistent learned dynamics into optimization-based underwater visual-inertial odometry, providing a principled framework for robust state estimation under environmental disturbances.

I. INTRODUCTION

Accurate state estimation is fundamental for autonomous robots operating in GPS-denied environments such as underwater domains, where turbulent fluids, erratic currents, and limited visibility degrade sensor data. Visual-inertial odometry (VIO) has become a standard approach for such environments, fusing camera and IMU measurements to estimate robot trajectories. However, conventional VIO treats the robot as a passive observer, ignoring actuation dynamics and absorbing unmodeled forces as measurement noise—a limitation that causes drift under environmental disturbances.

The challenge stems from a modeling gap: standard VIO systems predict motion purely from sensor observations, neglecting the robot's own dynamic model. In reality, observed motion results from the interplay between commanded actuation and environmental forces. When external forces are present, the resulting discrepancies bias visual-inertial residuals and degrade pose accuracy.

Recent work has explored dynamics-aware state estimation for aerial vehicles, where analytical models of thrust and drag can be incorporated into estimation pipelines. However, these analytical approaches require known physical parameters—mass, inertia tensors, drag coefficients—that are difficult to obtain for underwater vehicles. Hydrodynamic forces are highly nonlinear and environment-dependent; added mass, damping, and flow-induced effects vary with vehicle configuration, operating depth, and proximity to structures. This makes purely physics-based models impractical for general underwater deployment where parameter identification is costly or infeasible.

We target a minimal sensor configuration—monocular camera and IMU—common in low-cost inspection platforms where Doppler velocity logs (DVL) or multibeam sonar are prohibitively expensive, power-intensive, or impractical due to size constraints. This sensor-minimal setting increases the importance of exploiting all available information, including actuation commands, to constrain the estimation problem.

Physics-informed machine learning offers a path forward, by embedding physical constraints such as energy conservation into learned models, one can capture complex dynamics while maintaining physical consistency. Port-Hamiltonian Neural ODEs represent one such approach, learning system dynamics while enforcing passivity and energy dissipation. We leverage such a model to provide dynamics predictions without requiring prior knowledge of hydrodynamic parameters—the network learns inertial, damping, and actuation characteristics directly from recorded trajectories.

We present a visual-inertial odometry framework that integrates physically consistent learned dynamics into factor graph optimization for robust underwater state estimation. The key insight is that a learned dynamics model provides an independent prediction of control-induced motion, enabling explicit identification of environmental disturbances. By comparing this prediction with visual-inertial observations within a unified optimization, discrepancies can be attributed to an external force variable rather than corrupting pose estimates.

A notable property of the proposed formulation is robustness to model mismatch, if the dynamics prediction is inaccurate due to payload changes, actuator wear, or domain shift, the optimizer attributes the discrepancy to the external force variable rather than biasing pose estimates. This "leaking" of model error into force estimation preserves odometry accuracy even when the learned dynamics are imperfect which is critical for field deployment where operating conditions may differ.

The remainder of this paper is organized as follows: Section II reviews prior work on dynamics-aware state estimation and physics-informed learning. Section III presents

[1] Biomechatronics and Energy-Efficient Robotics Lab (BE2R Lab), ITMO University, Saint Petersburg, Russia. {yazanmaalla, s.kolyubin}@itmo.ru

the proposed methodology. Section IV reports experimental evaluation on real and simulated underwater data. Section V concludes with discussion of limitations and future directions.

II. RELATED WORK

A. Dynamics-Aware State Estimation

Early integration of dynamics into visual-inertial systems focused on aerial vehicles, where pre-integrated dynamics factors [1] extended IMU preintegration [2] to accumulate thrust and drag measurements within factor graphs. Building on this, VIMO [3] jointly optimized pose and external forces using analytical UAV dynamics, while VID-Fusion [4] improved force estimation accuracy at increased computational cost. Filter-based variants augment state estimators with dynamics models [5], and hybrid methods combine learning with estimation: DIDO [6] fuses CNN-based debiasing with an EKF, while HDVIO [7] couples temporal convolution networks with factor-graph optimization for aerodynamic force prediction. Although these methods demonstrate the benefit of dynamics-augmented estimation, they all target aerial platforms and rely on analytical dynamics models with known parameters, so their approaches do not directly transfer to the complex, parameter-uncertain hydrodynamics of underwater vehicles.

B. Physics-Informed Learning for Robot Dynamics

Physics-informed machine learning embeds physical laws into learned models to improve generalization. Physics-Informed Neural Networks [8] impose PDE constraints in loss functions, with UAV applications demonstrating improved state estimation [9], [10]. Structure-preserving architectures further enforce geometric and energetic invariants: Hamiltonian Neural Networks [11] conserve total energy, Deep Lagrangian Networks [12] embed Euler-Lagrange mechanics, and Port-Hamiltonian Neural ODEs [13] additionally capture dissipative forces while guaranteeing passivity. Building on this last family, recent work [14] applies Port-Hamiltonian Neural ODEs to underwater robots, achieving accurate long-horizon motion prediction while maintaining physical consistency. We leverage that pre-trained model to supply dynamics predictions within our estimation framework.

C. Underwater State Estimation with Dynamics

Model-aided underwater state estimation has received limited attention compared to aerial counterparts. Early work [15] fused simplified 3-DOF dynamics with current estimation in an EKF but suffered from linearization errors and required known model parameters, while pseudo-DVL methods [16] replace Doppler measurements with translational models but still neglect rotational and complex hydrodynamics. Learning-based approaches [17], [18] adapt dynamics online but lack physical consistency guarantees, and physics-informed methods [19] have been explored for control but not integrated into SLAM. Recent factor-graph systems [20] switch between VIO and kinematic models for robustness but still ignore hydrodynamic forces. DeepVL [21] regresses velocity from proprioceptive inputs using RNNs and fuses the learned estimates in an EKF; however, it does not perform force estimation or tightly couple dynamics into a factor graph. To our knowledge, no prior work has integrated physically consistent learned dynamics into underwater visual-inertial odometry.

D. External Force Estimation

Several approaches estimate external forces alongside state. VIMO [3] integrates analytical UAV dynamics into VIO for joint force sensing, while ViEW [22] addresses wrenches in legged robots via visual-inertial estimation. Filter-based methods augment the robot state with forces using UKFs for quadrotor wrenches [23] or MSCKFs with instantaneous accelerometer updates [24]. Domain-specific methods estimate ocean currents for AUVs [25] or wind via disturbance observers on UAVs [26], but most are platform-specific or rely on analytical dynamics with known parameters. Our approach provides dynamics-informed odometry with integrated force estimation using a learned model that requires no prior parameter knowledge of mass, inertia, or hydrodynamic coefficients.

III. METHODOLOGY

A. System Overview

We consider an underwater robot equipped with a monocular camera, IMU, and access to control commands. The goal is to estimate the robot's trajectory while remaining robust to external disturbances such as ocean currents and contact forces.

Let frames be denoted: w (world), b (body), c (camera). The pose of b in w is $(R_w^b, p_w^b) \in SE(3)$ with rotation $R_w^b \in SO(3)$ and translation $p_w^b \in \mathbb{R}^3$. Quaternion parameterization is q_w^b. Velocities are v_w^b (body in world), angular velocities ω_b (in body). We use $\otimes$ for quaternion multiplication and $\|\mathbf{e}\|_\Sigma^2 = \mathbf{e}^\top \Sigma^{-1} \mathbf{e}$ for Mahalanobis norm.

Conventional VIO predicts motion purely from sensor observations; unmodeled disturbances are absorbed as noise, biasing residuals and inducing drift. We instead couple control commands $\mathbf{u}_k$ with a learned, physically consistent dynamics model f_{dyn}, which predicts commanded acceleration $\mathbf{a}_k^{\text{d}} = f_{\text{dyn}}(\mathbf{x}_k, \mathbf{u}_k)$, then preintegrated into a disturbance-free pose increment $\Delta\mathbf{x}_{k,k+1}^{\text{cmd}}$. Visual-inertial cues simultaneously provide observed increment $\Delta\mathbf{x}_{k,k+1}^{\text{obs}}$; discrepancy reveals external disturbance f_e, introduced as an additional factor in the graph. Fig. 1 shows the full pipeline, which extends VINS-Mono [27]: the front-end extracts image features (reprojection factors) and preintegrates IMU data (inertial factors), while control inputs are processed by a Port-Hamiltonian Neural ODE model to generate dynamics priors. A dynamics residual enforces consistency between model-predicted and observed motion, allowing joint estimation of trajectory, biases, and external forces in a sliding-window factor graph. The output is the robot's trajectory with interpretable, time-varying force estimates.

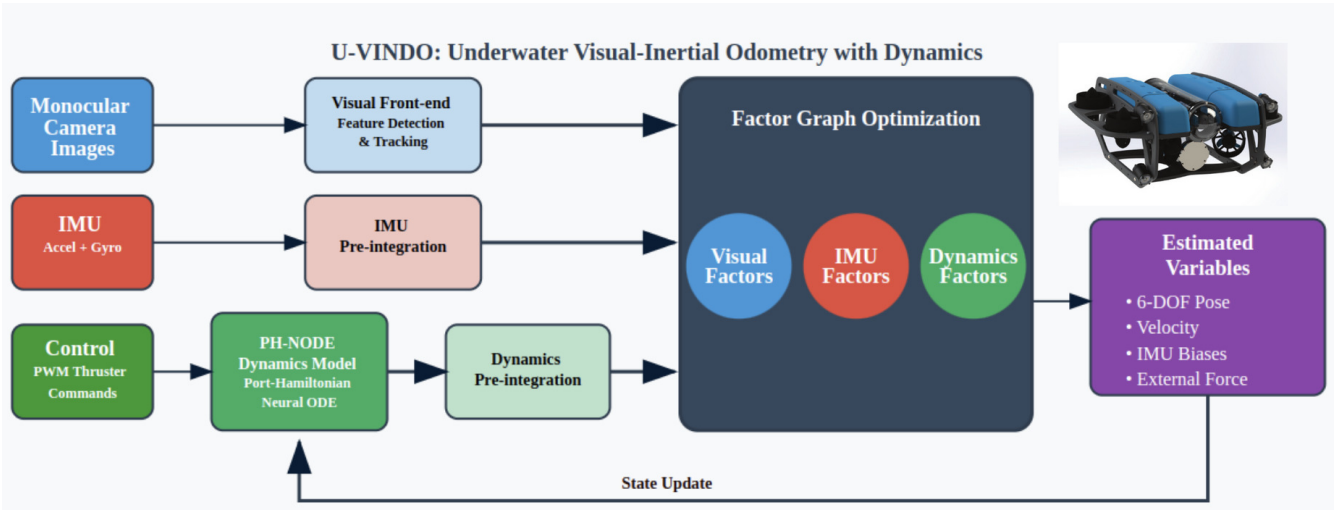


Fig. 1: System pipeline: Monocular features and IMU measurements form visual and inertial factors. Control commands (PWM) drive a learned dynamics model, preintegrated into dynamics factors. A sliding-window factor graph jointly estimates pose, velocity, IMU biases, and external force state.

B. State Representation

We augment the standard VIO state with an external disturbance term. At time t_k:

$$\mathbf{x}_k = \left[\mathbf{p}_k, \mathbf{q}_k, \mathbf{v}_k, \mathbf{b}_{g_k}, \mathbf{b}_{a_k}, f_{e_k}\right]^\top, \tag{1}$$

where $(\mathbf{p}_k, R(\mathbf{q}_k)) \in SE(3)$ is body pose in world frame, $\mathbf{v}_k$ is body linear velocity in world frame, (b_{g_k}, b_{a_k}) are IMU biases, and $f_{e_k} \in \mathbb{R}^3$ is external body-frame mass-normalized force. The sliding-window optimization variables are $\mathbf{X} = \{\mathbf{x}_k\}_{k=1}^n \cup \{\lambda_j\}_{j=1}^m$, with λ_j inverse-depth landmark parameters.

C. Factor Graph Formulation

Inspired by [3], we extend a visual-inertial back-end [27] with a dynamics factor. The joint objective is:

$$\mathcal{J} = e_{\text{visual}} + e_{\text{inertial}} + e_{\text{margin}} + e_{\text{dynamics}}, \tag{2}$$

where e_{visual} collects reprojection errors robustified with a Huber kernel, e_{inertial} encodes IMU preintegration constraints [2], e_{margin} carries priors from marginalized states, and $e_{\text{dynamics}} = \sum_{k=1}^n \|\mathbf{r}_k^d\|^2_{W_k^d}$ is the dynamics term detailed in Sec. III-E. Together these factors form a tightly coupled objective where classical VIO terms are augmented by dynamics-informed constraints and external-force estimation.

D. Learned Dynamics Model

The dynamics model must capture complex underwater hydrodynamics while maintaining physical consistency. We employ a Port-Hamiltonian Neural ODE (PH-NODE) formulation [14] that learns dynamics while enforcing energy conservation, to predict the body-frame acceleration that results from control inputs under nominal conditions (no external disturbance).

Let $\boldsymbol{\eta} \in SE(3)$ denote the robot pose, $\boldsymbol{\xi} \in \mathbb{R}^6$ its body-frame twist (linear and angular velocity), $\boldsymbol{\rho} \in \mathbb{R}^6$ the generalized momenta, and $\mathbf{u} \in \mathbb{R}^m$ the control input. The inertia matrix is denoted by $\mathbf{M}(\boldsymbol{\eta}) \in \mathbb{R}^{6\times6}$, the Hamiltonian by $H(\boldsymbol{\eta}, \boldsymbol{\rho})$, the dissipation matrix by $\boldsymbol{D}(\boldsymbol{\eta}, \boldsymbol{\rho})$, and $\boldsymbol{g}(\boldsymbol{\eta}, \mathbf{u})$ is a learnable actuation mapping. The Port-Hamiltonian neural dynamics can be expressed as:

$$\begin{aligned}
&\dot{\boldsymbol{\eta}} = f_{\boldsymbol{\eta}}(\boldsymbol{\eta}, \boldsymbol{\xi}) = \boldsymbol{\eta}\boldsymbol{\xi}^\wedge, \\
&\dot{\boldsymbol{\xi}} = f_{\boldsymbol{\xi}}(\boldsymbol{\eta}, \boldsymbol{\xi}, \mathbf{u}) = \mathbf{M}^{-1}(\boldsymbol{\eta})\dot{\boldsymbol{\rho}} + \dot{\mathbf{M}}^{-1}(\boldsymbol{\eta})\boldsymbol{\rho}, \\
\text{where}\quad &\boldsymbol{\rho} = \mathbf{M}(\boldsymbol{\eta})\boldsymbol{\xi}, \\
&\dot{\boldsymbol{\rho}} = \begin{bmatrix} \boldsymbol{\rho}_v \times \boldsymbol{\omega} - \mathbf{R}^\top \frac{\partial H(\boldsymbol{\eta},\boldsymbol{\rho})}{\partial \mathbf{p}} \\ \boldsymbol{\rho}_\omega \times \boldsymbol{\omega} + \boldsymbol{\rho}_v \times \mathbf{v} + \sum_{i=1}^3 \mathbf{r}_i \times \frac{\partial H(\boldsymbol{\eta},\boldsymbol{\rho})}{\partial \mathbf{r}_i} \end{bmatrix} \\
&\qquad - \boldsymbol{D}(\boldsymbol{\eta}, \boldsymbol{\rho})\boldsymbol{\xi} + \boldsymbol{g}(\boldsymbol{\eta}, \mathbf{u})
\end{aligned} \tag{3}$$

where $\boldsymbol{\eta}\boldsymbol{\xi}^\wedge$ encodes the standard kinematics on $SE(3)$, and the dynamics evolve according to the port-Hamiltonian structure with dissipation and actuation terms.

While the kinematic model $f_{\boldsymbol{\eta}}$ is fully specified, the dynamic model $f_{\boldsymbol{\xi}}$ depends on system parameters that are not known a priori. Specifically, the inertia matrix $\mathbf{M}(\boldsymbol{\eta})$, the potential energy $V(\boldsymbol{\eta})$, the dissipation matrix $\boldsymbol{D}(\boldsymbol{\eta}, \boldsymbol{\rho})$, and the actuation model $\boldsymbol{g}(\boldsymbol{\eta}, \mathbf{u})$ are all approximated by neural networks:

$$\mathbf{M}(\boldsymbol{\eta}; \boldsymbol{\theta}_M), \quad V(\boldsymbol{\eta}; \boldsymbol{\theta}_V), \quad \boldsymbol{D}(\boldsymbol{\eta}, \boldsymbol{\rho}; \boldsymbol{\theta}_D), \quad \boldsymbol{g}(\boldsymbol{\eta}, \mathbf{u}; \boldsymbol{\theta}_g).$$

To ensure physical consistency, the mass and dissipation matrices are parameterized via Cholesky decomposition, guaranteeing symmetry and positive definiteness. Here, $\boldsymbol{\theta}_M$, $\boldsymbol{\theta}_V$, $\boldsymbol{\theta}_D$, and $\boldsymbol{\theta}_g$ denote the learnable parameter vectors of each respective neural network component. In practice, each of $\mathbf{M}$, V, $\mathbf{D}$, and $\mathbf{g}$ is realized as a two-layer MLP with 256 hidden units and Tanh activations; the mass and dissipation

sub-networks are further split into linear and angular parts as in (3).

Training: The PH-NODE is trained following the procedure described in [14], using the NTNU underwater dataset with thruster PWM commands and ground-truth poses as inputs. Since the dataset provides pose only, body-frame twist is approximated via a fifth-order forward-backward Butterworth filter followed by center-point numerical differentiation. The training objective combines a geodesic distance loss on $SE(3)$ accumulated over the full prediction horizon with an ℓ_1 sparsity penalty on $\boldsymbol{\theta}_g$ to prevent the actuation network from absorbing physical structure. Full architecture details, hyperparameters, and dataset splits are reported in [14].

Transferability: Because $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{g}$ are learned from data, deploying U-VINDO on a vehicle with different physical characteristics requires retraining the PH-NODE on data from that platform. No system-identification experiments or prior knowledge of mass, inertia, or drag coefficients are needed: training requires only recorded trajectories with state and control inputs under nominal conditions. This makes adaptation straightforward, at the cost of a new data collection campaign.

The first three components of $\dot{\boldsymbol{\xi}}$ yield the commanded body-frame linear acceleration $\mathbf{a}_d^b$ due to the control inputs (with no external disturbances). In other words, $\mathbf{a}_d^b$ is the acceleration that the robot's dynamics would produce from the thruster commands alone, under normal operating conditions.

In the proposed dynamics-augmented framework, we preintegrate $\mathbf{a}_d^b$ between keyframes to produce a disturbance-free motion prior, and define a dynamic residual r^d so that discrepancies between this prior and the estimated motion are explained by an explicit external force f_e optimized within the graph. This yields interpretable, time-varying force estimates while preserving the benefits of tightly coupled visual–inertial optimization. For more details on the Port-Hamiltonian mechanics on manifolds, we refer the reader to [13], and for adaptation to underwater we refer to [14].

E. Dynamics Preintegration

The dynamics factor enables external-force estimation by comparing motion predicted by the learned model with motion observed from visual-inertial odometry. The model outputs body-frame linear acceleration $\mathbf{a}_d^b$ induced by control commands, trained on data collected under minimal external disturbances so that $\mathbf{a}_d^b$ reflects nominal actuation-driven motion. External interactions (currents, contacts) are represented as an additional body-frame acceleration f_e^b.

To handle high-rate control inputs efficiently, we preintegrate the dynamics between keyframes, analogously to IMU preintegration [2]. The continuous-time kinematics in the world frame are:

$$\dot{\mathbf{p}}_b^w = \mathbf{v}_b^w, \qquad \dot{\mathbf{v}}_b^w = \mathbf{R}_b^w(\mathbf{a}_d^b + f_e^b), \tag{4}$$

which separates commanded motion from unmodeled disturbances. We place a zero-mean Gaussian prior on the external acceleration, $f_e^b \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_f)$, reflecting that disturbances are typically small unless data suggest otherwise.

Integrating over $[t_k, t_{k+1}]$ and assuming f_e^b is constant over the interval, we obtain predicted increments (from optimized states) and observed increments (from preintegrated dynamics):

Predicted increments (state-based), transforming world-frame estimates to body frame with $\Delta t = t_{k+1} - t_k$:

$$\boldsymbol{\alpha}_k^{k+1} = \mathbf{R}_w^{b_k}(\mathbf{p}_{k+1} - \mathbf{p}_k - \mathbf{v}_k\Delta t) - \tfrac{1}{2}f_{e_k}\Delta t^2, \tag{5}$$

$$\boldsymbol{\beta}_k^{k+1} = \mathbf{R}_w^{b_k}(\mathbf{v}_{k+1} - \mathbf{v}_k) - f_{e_k}\Delta t. \tag{6}$$

Observed increments (measurement-based), preintegrating the learned acceleration, where $\hat{\boldsymbol{\alpha}}_k^{k+1}$ and $\hat{\boldsymbol{\beta}}_k^{k+1}$ denote the preintegrated position and velocity increments respectively:

$$\hat{\boldsymbol{\alpha}}_k^{k+1} = \int_{t_k}^{t_{k+1}} \int_{t_k}^{\tau} \mathbf{R}_{b_\sigma}^{b_k}\mathbf{a}_d^{b_\sigma}\, d\sigma\, d\tau, \tag{7}$$

$$\hat{\boldsymbol{\beta}}_k^{k+1} = \int_{t_k}^{t_{k+1}} \mathbf{R}_{b_\tau}^{b_k}\mathbf{a}_d^{b_\tau}\, d\tau. \tag{8}$$

In practice, assuming $\mathbf{a}_d^b$ is constant over each small integration step δt, we recursively update:

$$\hat{\boldsymbol{\gamma}}_{i+1}^{b_k} = \hat{\boldsymbol{\gamma}}_i^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2}(\tilde{\boldsymbol{\omega}}_i - \mathbf{b}_{g_i})\delta t \end{bmatrix}, \tag{9}$$

$$\hat{\boldsymbol{\beta}}_{i+1}^{b_k} = \hat{\boldsymbol{\beta}}_i^{b_k} + R(\hat{\boldsymbol{\gamma}}_i^{b_k})\mathbf{a}_d^{b_i}\delta t, \tag{10}$$

$$\hat{\boldsymbol{\alpha}}_{i+1}^{b_k} = \hat{\boldsymbol{\alpha}}_i^{b_k} + \hat{\boldsymbol{\beta}}_i^{b_k}\delta t + \tfrac{1}{2}R(\hat{\boldsymbol{\gamma}}_i^{b_k})\mathbf{a}_d^{b_i}\delta t^2, \tag{11}$$

where $\hat{\boldsymbol{\gamma}}$ denotes the orientation increment from IMU measurements $\tilde{\boldsymbol{\omega}}$.

F. Dynamics Residual

The dynamics residual penalizes disagreement between state-predicted increments and dynamics-observed increments, together with a zero-mean prior on f_e:

$$\mathbf{r}_d = \begin{bmatrix} \boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}} \\ \boldsymbol{\beta} - \hat{\boldsymbol{\beta}} \\ f_e \end{bmatrix}, \quad \mathbf{W}_d^k = \mathrm{diag}(\mathbf{P}_{\alpha\beta}^{-1}, \boldsymbol{\Sigma}_f^{-1}), \tag{12}$$

where $\mathbf{P}_{\alpha\beta}$ is the covariance of $(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}})$ from preintegration.

The residual couples f_e to the trajectory so that discrepancies may be explained either by state corrections or by nonzero f_e. The first two blocks enforce that observed motion equals commanded motion plus f_e; thus, if the trajectory already explains the data, f_e remains near zero. Otherwise, persistent mismatches drive f_e away from zero, indicating genuine external disturbances. The zero-mean prior serves as Bayesian regularization preventing overfitting of small errors as forces.

G. Dynamics Covariance Estimation

Proper weighting of dynamics factors is critical: overconfidence overwhelms sensor data while underconfidence diminishes benefit. We estimate covariance empirically on

TABLE I: Real-World Evaluation (NTNU Underwater Dataset). APE: Absolute Position Error RMSE [m], ARE: Absolute Rotation Error RMSE [deg]. Best results in **bold**.

Sequence	Length [m]	APE [m]		ARE [deg]	
		Baseline	Ours	Baseline	Ours
fjord_1	156.6	**0.428**	0.445	1.233	**1.072**
fjord_2	287.3	**1.129**	1.142	**2.654**	3.067
fjord_3	183.0	**0.776**	0.789	**1.963**	2.055
fjord_4	229.9	**0.393**	0.490	**1.235**	1.341
fjord_5	239.5	0.851	**0.832**	1.863	**1.540**
fjord_6	387.4	1.245	**1.211**	2.872	**2.361**
mclab_1	133.8	0.594	**0.417**	1.842	**1.513**
mclab_2	128.8	0.367	**0.340**	1.984	**1.713**
Average	–	0.848	**0.833**	1.956	**1.833**

validation data. Given predictions $\mathbf{a}_{d,k} = f_{\text{dyn}}(\mathbf{x}_k, \mathbf{u}_k)$ and ground truth accelerations $\mathbf{a}_{gt,k}$:

$$\boldsymbol{\epsilon}_k = \mathbf{a}_{d,k} - \mathbf{a}_{gt,k}, \quad \boldsymbol{\Sigma}_{\text{dyn}} = \frac{1}{N-1}\sum_{k=1}^{N}(\boldsymbol{\epsilon}_k - \bar{\boldsymbol{\epsilon}})(\boldsymbol{\epsilon}_k - \bar{\boldsymbol{\epsilon}})^\top. \tag{13}$$

This covariance is propagated through integration to weight the dynamics factor appropriately.

IV. EXPERIMENTAL EVALUATION

A. Real-World Evaluation

We evaluate on the NTNU underwater dataset, comprising sequences in natural environment (fjord) and laboratory (mclab) settings with ground truth. Table I compares our method against VINS-Mono baseline.

On average, the proposed method reduces APE by 1.8% and ARE by 6.3%. While improvements are modest in these sequences—which lack strong persistent disturbances—the dynamics factor acts as a stabilizing prior without degrading performance. Importantly, the dynamics factor adds only 0.6 ms average processing time per frame (baseline: 19.7 ms $\rightarrow$ ours: 20.3 ms), representing less than 4% computational overhead.

Figure 2 shows trajectory and error comparisons for the fjord_5 sequence, where the dynamics-informed approach provides consistent improvements in both position and orientation.

B. Simulation with Controlled Disturbances

To evaluate robustness under known disturbances, we use the HoloOcean underwater simulator with controlled force injection. External forces of varying magnitudes and directions are applied at predefined intervals along trajectories.

Table II shows substantial improvements: average APE reduction of 61% and ARE reduction of 49%. Figure 3 illustrates trajectory comparisons, where baseline VIO drifts during force injection while the proposed method maintains accuracy.

TABLE II: Simulation Evaluation with Known Disturbances. APE: Absolute Position Error RMSE [m], ARE: Absolute Rotation Error RMSE [deg].

Sequence	Length [m]	APE [m]		ARE [deg]	
		Baseline	Ours	Baseline	Ours
Seq. 1	446.8	1.519	**0.844**	1.523	**0.716**
Seq. 2	192.5	1.183	**0.631**	0.877	**0.453**
Seq. 3	610.8	1.312	**0.579**	0.572	**0.387**
Seq. 4	646.6	1.261	**0.564**	1.090	**0.591**
Seq. 5	557.8	1.764	**0.362**	0.676	**0.385**
Seq. 6	442.8	2.292	**0.643**	1.612	**0.713**
Average	–	1.555	**0.604**	1.058	**0.541**

TABLE III: Force Estimation Accuracy Across Simulation Sequences

Sequence	RMSE [N/kg]	MAE [N/kg]	Correlation
Seq. 1	0.797	0.765	0.984
Seq. 2	0.292	0.268	0.995
Seq. 3	0.802	0.770	0.984
Seq. 4	0.781	0.734	0.981
Seq. 5	0.746	0.714	0.985
Seq. 6	0.765	0.732	0.984
Average	**0.697**	**0.664**	**0.986**

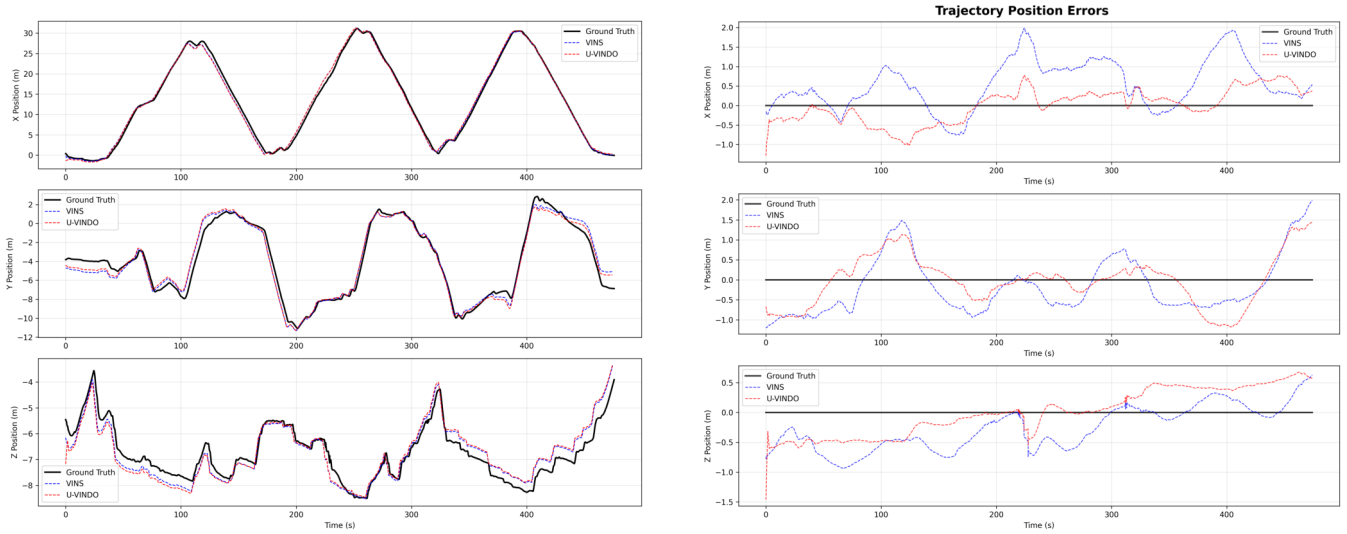
C. External Force Estimation

Beyond trajectory accuracy, the framework provides interpretable force estimates suitable for higher-level autonomy applications. Figure 4 shows force estimation results: both amplitude and timing are accurately recovered. Table III summarizes performance across sequences: mean correlation exceeds 0.98 with average errors below 0.7 N/kg. These results confirm that the optimizer consistently attributes unmodeled dynamics to the external force variable rather than biasing pose estimates.

The estimated forces provide actionable information beyond odometry: detecting contact events during inspection or manipulation tasks, estimating ambient current fields for energy-efficient trajectory planning, identifying potential tether entanglement, or triggering mission re-planning when environmental disturbances approach or exceed the vehicle's control authority.

D. Discussion

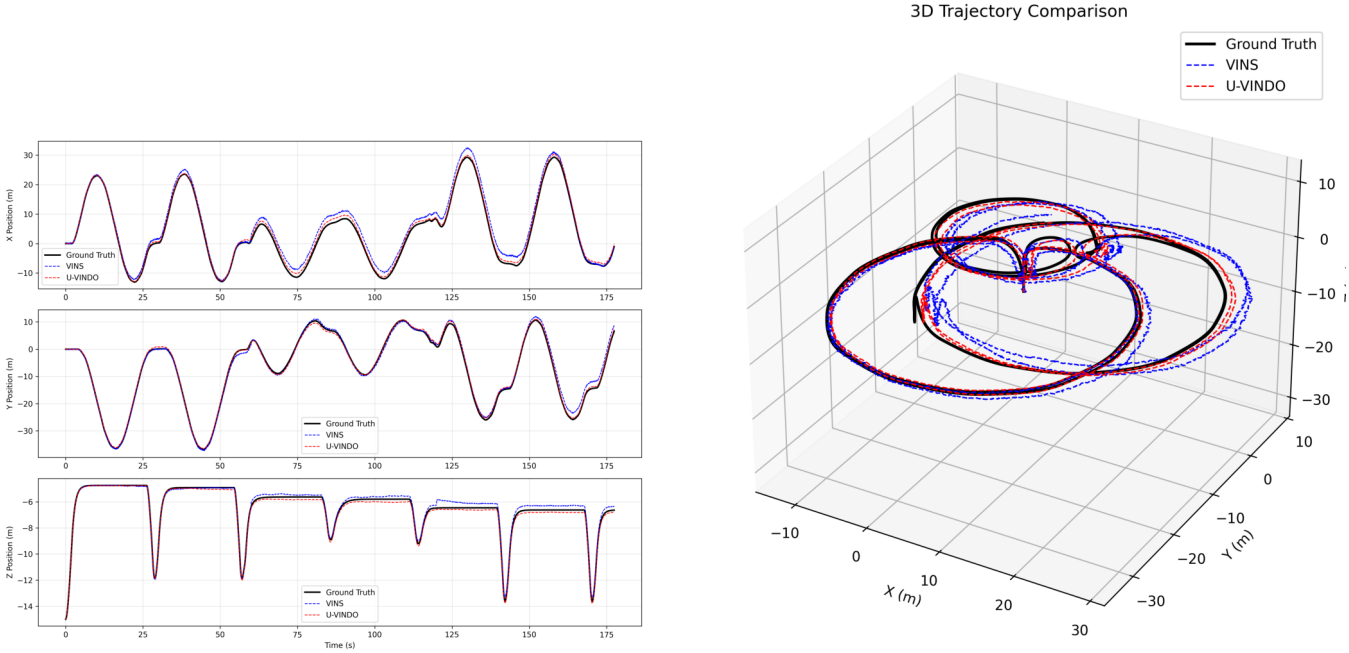
The proposed framework most closely parallels VIMO [3] in its factor graph structure and force estimation formulation, but differs in two fundamental respects: VIMO requires analytical UAV dynamics with known parameters, while U-VINDO uses a learned, parameter-free model; and VIMO targets aerial platforms with relatively simple aerodynamics, whereas we address the substantially more complex hydrodynamic regime of underwater vehicles. Compared to HDVIO [7], which uses temporal convolution networks for aerodynamic prediction, our Port-Hamiltonian structure explicitly enforces passivity and energy conservation, providing stronger physical guarantees. DeepVL [21], the closest underwater counterpart, does not estimate external forces and



(a) Estimated trajectories.

(b) Position errors over time.

Fig. 2: Real-world evaluation (NTNU dataset, fjord_5 sequence): (a) Trajectory comparison showing baseline VIO (blue dashed), proposed method (red dashed), and ground truth (black solid). (b) Per-axis position error signals demonstrating consistent reduction across all components.



(a) Position over time.

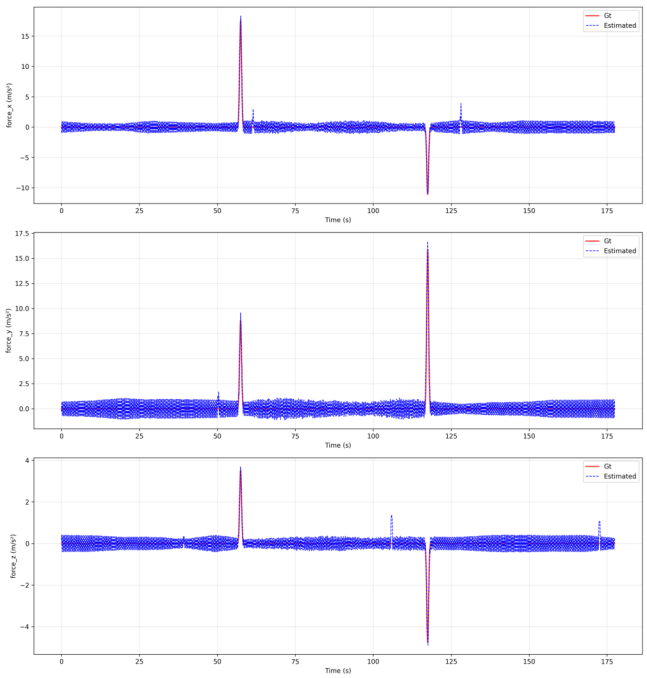
(b) 3D trajectory view.

Fig. 3: Simulation evaluation with controlled force disturbances: Baseline VIO (blue) shows significant drift during force injection periods, while the proposed method (red) maintains accuracy close to ground truth (black).

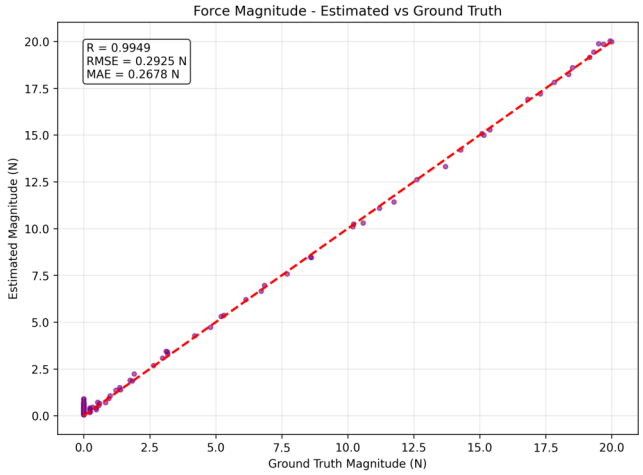
relies on an EKF rather than a tightly coupled factor graph; our method provides richer state estimates at comparable sensor overhead.

The experiments demonstrate that learned dynamics priors improve pose estimation when environmental disturbances are present while remaining neutral in benign conditions. On real data without strong disturbances, gains are modest because baseline VIO already explains most motion; nevertheless, the dynamics factor provides stabilization at minimal computational cost.

Under disturbances, the dynamics residual enables principled attribution of discrepancies to external forces, preventing drift in visual-inertial residuals. A key property of the formulation is graceful degradation: if the dynamics prediction is inaccurate due to model mismatch—for instance, from payload changes, actuator degradation, or



(a) Force components over time.



(b) Estimated vs. true force.

Fig. 4: Force estimation results: (a) Estimated (blue) vs. ground truth (red) force components, accurately recovering both amplitude and timing of injected disturbances. (b) Regression analysis across all simulation sequences showing near-perfect correlation ($R = 0.99$).

operating conditions differing from training—the optimizer assigns the discrepancy to the force variable rather than corrupting pose estimates. This leaking of model error into force estimation preserves odometry accuracy even when the learned dynamics are imperfect, a critical property for field deployment.

The current formulation uses a zero-mean prior on the external force state, which performs well for transient disturbances as demonstrated in our experiments. However, this choice may lead to degraded performance under strong persistent forces such as sustained ocean currents, where the zero-mean assumption conflicts with the true force profile. Addressing this limitation through alternative prior structures—such as random-walk priors for persistent components—is an important direction for future work.

Additional limitations include the assumption of constant forces over preintegration intervals, reliance on accurate control signals, and focus on translational dynamics (the current factor does not directly constrain orientation). Extending to full 6-DoF disturbance estimation and incorporating online adaptive covariance are promising directions for enhanced robustness.

V. CONCLUSION

We presented a dynamics-informed visual-inertial odometry framework for robust underwater state estimation. By integrating physically consistent learned dynamics into a tightly coupled factor graph via a preintegration scheme, the system explicitly separates commanded motion from environmental disturbances. The approach requires no prior knowledge of hydrodynamic parameters, operating directly from raw thruster commands—a significant advantage for field deployment where parameter identification is impractical.

Evaluation on real underwater data demonstrates consistent accuracy improvements with ignorable computational overhead. In controlled simulations with known disturbances, the framework reduces trajectory error and achieves near-perfect force estimation correlation. The estimated forces provide actionable information for higher-level autonomy tasks including contact detection, current estimation, and adaptive mission planning.

To our knowledge, this represents the first integration of physically consistent learned dynamics into optimization-based underwater visual-inertial odometry.

REFERENCES

[1] A. Antonini, "Pre-integrated dynamics factors and a dynamical agile visual-inertial dataset for uav perception," Ph.D. dissertation, Massachusetts Institute of Technology, 2018.

[2] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Transactions on Robotics*, vol. 33, pp. 1–21, 2015.

[3] B. Nisar, P. Foehn, D. Falanga, and D. Scaramuzza, "Vimo: Simultaneous visual inertial model-based odometry and force estimation," *IEEE Robotics and Automation Letters*, vol. 4, no. 3, pp. 2785–2792, 2019.

[4] Z. Ding, T. Yang, K. Zhang, C. Xu, and F. Gao, "Vid-fusion: Robust visual-inertial-dynamics odometry for accurate external force estimation," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 14 469–14 475, 2021.

[5] O. Omotuyi and M. Kumar, "Uav visual-inertial dynamics (vi-d) odometry using unscented kalman filter," *IFAC PapersOnLine*, vol. 54, no. 20, pp. 814–819, 2021.

[6] K. Zhang, C. Jiang, J. Li, *et al.*, "Dido: Deep inertial quadrotor dynamical odometry," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 9083–9090, 2022.

[7] G. Cioffi, L. Bauersfeld, and D. Scaramuzza, "Hdvio: Improving localization and disturbance estimation with hybrid dynamics vio," *ArXiv*, vol. abs/2306.11429, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259203584.

[8] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, 2019.

[9] D. Bianchi, N. Epicoco, M. D. Ferdinando, S. D. Gennaro, and P. Pepe, "Physics-informed neural networks for unmanned aerial vehicle system estimation," *Drones*, vol. 8, no. 716, 2024, ISSN: 2504-446X.

[10] W. Gu, S. Primatesta, and A. Rizzo, "Physics-informed neural network for quadrotor dynamical modeling," *Robotics and Autonomous Systems*, vol. 171, p. 104 569, 2024.

[11] S. Greydanus, M. Dzamba, and J. Yosinski, "Hamiltonian neural networks," 2019. arXiv: 1906.01563 [cs.NE].

[12] M. Lutter, C. Ritter, and J. Peters, "Deep lagrangian networks: Using physics as model prior for deep learning," in *International Conference on Learning Representations (ICLR)*, 2019. arXiv: 1907.04490 [cs.LG].

[13] T. Duong, A. Altawaitan, J. Stanley, and N. Atanasov, "Port-hamiltonian neural ODE networks on lie groups for robot dynamics learning and control," *IEEE Transactions on Robotics*, vol. 40, pp. 3695–3715, 2024.

[14] Z. A. Barhoum and S. Kolyubin, "Physically consistent dynamic modeling of underwater robots for robust long-horizon motion prediction," *Journal of Instrument Engineering*, vol. 68, no. 11, pp. 983–995, 2025.

[15] A. Martinez, L. Hernandez, H. Sahli, Y. Valeriano-Medina, M. Orozco-Monteagudo, and D. Garcia-Garcia, "Model-aided navigation with sea current estimation for an autonomous underwater vehicle," *International Journal of Advanced Robotic Systems*, vol. 12, p. 103, 2015.

[16] A. Karmozdi, M. Hashemi, H. Salarieh, and A. Alasty, "Implementation of translational motion dynamics for ins data fusion in dvl outage in underwater navigation," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6652–6662, 2021.

[17] B. Wehbe, M. Hildebrandt, and F. Kirchner, "A framework for on-line learning of underwater vehicles dynamic models," *2019 International Conference on Robotics and Automation (ICRA)*, pp. 7969–7975, 2019.

[18] X. Macatangay, S. A. Gabriel, R. Hoseinnezhad, A. Fowler, and A. Bab-Hadiashar, "Machine learning for modeling underwater vehicle dynamics: Overview and insights," *IEEE Access*, vol. 12, pp. 139 486–139 504, 2024. DOI: 10.1109/ACCESS.2024.3464644.

[19] Y. Zhao, Z. Hu, W. Du, L. Geng, and Y. Yang, "Research on modeling method of autonomous underwater vehicle based on a physics-informed neural network," *Journal of Marine Science and Engineering*, vol. 12, no. 801, 2024.

[20] B. Joshi, H. Damron, S. Rahman, and I. Rekleitis, "Sm/vio: Robust underwater state estimation - switching between model-based and visual inertial odometry," in *IEEE Conference on Robotics and Automation (ICRA)*, 2023.

[21] M. Singh and K. Alexis, "Deepvl: Dynamics and inertial measurements-based deep velocity learning for underwater odometry," 2025. arXiv: 2502.07726 [cs.RO].

[22] J. Kang, H. Kim, and K.-S. Kim, "View: Visual-inertial external wrench estimator for legged robot," *IEEE Robotics and Automation Letters*, vol. 8, no. 12, pp. 8366–8373, 2023. DOI: 10.1109/LRA.2023.3322646.

[23] M. McKinnon and A. P. Schoellig, "State estimation for aggressive flight in cluttered environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[24] J. Song, A. Richard, and M. Olivares-Mendez, *An accurate filter-based visual inertial external force estimator via instantaneous accelerometer update*, 2024. arXiv: 2408.16354 [cs.RO].

[25] B. Osborn, "Autonomous underwater vehicle navigation using current estimation and dead reckoning," M.S. thesis, University of Idaho, USA, 2021.

[26] H. Yu, X. Liang, and X. Lyu, "Dob-based wind estimation of a uav using its onboard sensor," in *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024, pp. 8126–8133.

[27] T. Qin, P. Li, and S. Shen, "Vins-mono: A robust and versatile monocular visual-inertial state estimator," *IEEE Transactions on Robotics*, vol. 34, pp. 1004–1020, 2017.