Integrated Stereo Vision and Compliance Control With an Underwater Manipulator for Hydraulic Structure Inspection and Maintenance

Published Version:
Wang, F., Liu, H., Cao, L., Song, G., Song, A., Meattini, R., et al. (2026). Integrated Stereo Vision and Compliance Control With an Underwater Manipulator for Hydraulic Structure Inspection and Maintenance. IEEE/ASME TRANSACTIONS ON MECHATRONICS, 1, 1-12 [10.1109/tmech.2026.3677878].

Availability:
This version is available at: https://hdl.handle.net/11585/1063953 since: 2026-05-18

Published:
DOI: http://doi.org/10.1109/tmech.2026.3677878

(Article begins on next page)

Integrated Stereo Vision and Compliance Control with an Underwater Manipulator for Hydraulic Structure Inspection and Maintenance

Fei Wang, Haixin Liu, Lin Cao, Guangming Song, *Senior Member, IEEE*, Aiguo Song, *Senior Member, IEEE*, Roberto Meattini, *Member, IEEE*, and Claudio Melchiorri, *Life Fellow, IEEE*

***Abstract*—Hydraulic structure inspection and maintenance are essential for structural health and operational safety. To address the limitations of underwater intervention, this article introduces an integrated underwater manipulator system that supports both teleoperation and automatic operation modes. The automatic mode incorporates visual position estimation for locating structures, robust motion capability for accurate trajectory tracking, and compliance and force tracking for safe interaction. An uncertainty-based underwater disparity estimation network (UWNet) is developed to enhance position estimation accuracy. This network combines global and local features effectively and integrates uncertainty estimation and pseudo-label generation to adapt the pre-trained UWNet, improving underwater disparity predictions. Additionally, a robust controller is designed, comprising an inner-loop position controller based on modified unknown system dynamics estimator (USDE) and super-twisting sliding mode control (ST SMC) to mitigate hydrodynamic disturbances and model uncertainties. An outer-loop variable admittance controller with adaptive feedback compensation ensures compliant interaction and force tracking. Experimental results show that, with reliable disparity estimation from the adapted UWNet, the system achieves a trajectory tracking RMSE of 1.44 mm at 1.0 m/s flow and a mean force error of 0.63 N during thickness measurement, while enabling stable surface cleaning under varying flow conditions, thereby facilitating hydraulic structure inspection and maintenance.**

***Index Terms*—Underwater manipulator, underwater disparity estimation, variable admittance control, hydraulic structure maintenance.**

I. Introduction

HYDRAULIC structures—such as pipelines, bridges, and dams—are constantly growing both in number and in service life, making the assurance of their structural integrity and operational safety increasingly critical. Currently, underwater inspection and maintenance are predominantly performed by human divers, whose operations are limited to shallow water environments and involve significant safety risks, leading to a growing demand for underwater robotic systems to replace human divers [1]. Although remotely operated vehicles (ROVs) and autonomous underwater vehicles (AUVs) are effective for visual inspections, they are unsuitable for contact-based inspection and maintenance tasks [2]. For such operations, underwater manipulators equipped with robust control algorithms and specialized tools offer a promising solution. Advanced robotic platforms such as underwater vehicle manipulator systems (UVMS) [3], crawler-type underwater robots [4], and underwater legged robots (ULR) [5] integrated with underwater manipulators are emerging as ideal candidates for performing contact-based underwater operations.

Contact-based inspection and maintenance of hydraulic structures involve a range of underwater intervention tasks, including object grasping, valve operation, switch toggling, pipeline inspection, and surface cleaning [6]. To perform these tasks safely and effectively, an underwater vision system is essential for accurately determining the position of the structures. Additionally, a robust manipulator control algorithm is required to ensure precise trajectory tracking and compliant interaction during task execution.

Underwater vision systems provide visual feedback to operators, enabling teleoperation, which remains a fundamental approach for underwater intervention tasks [7]. By enhancing the capability of remote manipulators to interact with unknown environments [8] or by adopting human-robot shared control strategies [9], [10], operators can perform complex tasks more safely and efficiently. To further improve efficiency and automation, increasing research has focused on utilizing visual information for pose estimation and autonomous robotic intervention [11]. Early studies demonstrated that tasks such as grasping [12] and valve turning [13] primarily relied on prior knowledge of target object features. Artificial markers have also been widely employed to facilitate interaction tasks, such as connector plugging/unplugging [14] and contact inspection [15]. Recently, marker-free deep learning techniques have been increasingly applied to underwater robotic intervention, achieving significant progress in grasping applications [16]. However, for hydraulic structure inspection and cleaning tasks, autonomous methods based on deep learning remain underex-

This work was supported in part by the National Natural Science Foundation of China under Grant 52127813 and Grant 62373098, and in part by the China Scholarship Council under Grant 202406090238 (Corresponding author: Guangming Song.)

Fei Wang is with the State Key Laboratory of Digital Medical Engineering, Jiangsu Key Lab of Robot Sensing and Control, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China, and also with the Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Bologna 40136, Italy (e-mail: neilwang@seu.edu.cn).

Haixin Liu, Lin Cao, Guangming Song, and Aiguo Song are with the State Key Laboratory of Digital Medical Engineering, Jiangsu Key Lab of Robot Sensing and Control, School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China (e-mail: lhaixin@seu.edu.cn; 220233619@seu.edu.cn; mikesong@seu.edu.cn; a.g.song@seu.edu.cn).

Roberto Meattini and Claudio Melchiorri are with the Department of Electrical, Electronic and Information Engineering (DEI), University of Bologna, Bologna 40136, Italy (e-mail: roberto.meattini@unibo.it; claudio.melchiorri@unibo.it).

plored, with teleoperation still dominant. The lack of underwater disparity datasets and the large domain shift between terrestrial and underwater images hinder the application of state-of-the-art disparity estimation methods, such as GwcNet [17], CFNet [18], and their subsequent improvements, in underwater environments. UCFNet [19] introduces an uncertainty-based domain adaptation framework that use uncertainty estimation to generate pseudo-labels, enabling pre-trained models to adapt to target domains and improve disparity estimation accuracy. This approach offers a promising basis for underwater vision-based localization of hydraulic structures.

Acquiring the positional information of hydraulic structures is crucial for successfully executing diverse underwater intervention tasks. However, due to hydrodynamic effects, trajectory tracking for underwater robots remains challenging [20]. Robust control strategies based on disturbance observer (DOB) and sliding mode controller (SMC) are widely employed to enhance trajectory tracking performance [21], [22]. More advanced uncertainty compensation and disturbance rejection methods, such as unknown system dynamics estimator (USDE) [23], [24] and super-twisting sliding mode control (ST SMC) [25], have shown effectiveness in collaborative robots but remain largely unvalidated in underwater environments. When an underwater manipulator performs contact inspection and maintenance, inaccuracies in visual position estimation and trajectory tracking can potentially damage both the manipulator and the hydraulic structure, making compliance control and force control essential for safeguarding both systems [26]. Patryk et al. [15] implemented adaptive admittance control by modifying stiffness parameters, enabling contact force tracking during pipe inspection, albeit with noticeable force overshoot and fluctuation. DexROV performed remotely operated intervention tasks in an oil-gas industry scenario, integrating admittance control to ensure compliance against undesired external forces [27]. In [28], a hybrid force-position control enhanced by SMC achieved trajectory and contact force tracking but exhibited significant chattering in contact force and control torque. These force control methods present certain limitations, including insufficient performance, testing only in still water, and contact restricted to flat surfaces.

Considering the requirements for hydraulic structure inspection and maintenance, as well as the limitations of underwater manipulators in position estimation, precise motion, and safe interaction, this article proposes a novel underwater manipulator system supporting both teleoperation and automatic operation modes. The automatic mode features visual position estimation for locating hydraulic structures, robust motion capability to ensure accurate trajectory tracking from non-contact to contact spaces, and compliance and force tracking to satisfy the safety demands during underwater inspection and maintenance. Compared with existing research, the main contributions of this article are as follows:

1) A comprehensive underwater electric manipulator system, featuring position estimation, precise motion, and compliant interaction, is developed. The system supports a teleoperation mode for complex tasks and an automatic mode for specific inspection and maintenance tasks.
2) An uncertainty-based underwater disparity estimation network (UWNet) is proposed, building on GwcNet by refining the combined volume module and integrating the uncertainty estimation and pseudo-label generation from UCFNet to adapt the pre-trained UWNet, thereby enhancing underwater disparity estimation accuracy.
3) A robust controller is designed, consisting of an inner-loop position controller based on modified USDE and ST SMC to counteract hydrodynamic disturbances and model uncertainties, and an outer-loop variable admittance controller with adaptive feedback compensation for compliant interaction and precise force tracking.

II. System Design

The overall structure of the developed underwater manipulator system and its operation process are illustrated in Fig. 1. The electric manipulator has a 6-degree-of-freedom (6-DoF) design, offering versatility for various operational poses (see Fig. 1(c)). Detailed internal specifications are provided in our previous study [29]. An underwater 6-axis force/torque (F/T) sensor is integrated on the wrist to measure contact forces between the end-effector and hydraulic structures. The end-effector for contact inspection is an ultrasonic thickness gauge, which can be interchanged with other functional probes. For surface cleaning tasks, the end-effector can be replaced with a rotary brush. An underwater binocular camera on the wrist allows operators to observe the underwater environment and choose the appropriate operating mode (see Fig. 1(a)).

To accommodate different conditions, two operating modes are employed, as depicted in Fig. 1(b) and implemented using Robot Operating System 2 (ROS2). In turbid water or complex tasks, the operator fully controls the manipulator via a haptic device in teleoperation mode. A traditional admittance controller ensures compliance and safety during underwater operations. In clear water and specific tasks, automatic operation mode is used. The developed disparity estimation network estimates the position of hydraulic structures and generates a stereo point cloud, which is displayed in the RViz interactive interface. The operator determines whether to execute point-contact tasks (e.g., inspecting a specific point) or continuous-path contact tasks (e.g., surface cleaning) and selects the starting point on the hydraulic structure. Based on the point cloud, task type, and starting point, the system automatically plans the path and generates the desired trajectory. A newly designed variable admittance controller not only provides compliance but also tracks the desired force during operation. Meanwhile, the inner-loop robust position controller precisely follows the desired joint angles generated by admittance controllers, allowing the manipulator to achieve high-accuracy execution.

This underwater manipulator system can be integrated into various underwater robotic platforms such as crawler, ROV, and ULR, for hydraulic structure inspection and maintenance. Underwater robots equipped with manipulators typically operate in a fixed-base mode (e.g., crawlers, ULRs) or a floating-base mode (e.g., ROVs), although ROVs can transition to fixed-base mode by landing on the seabed [3]. The control strategies presented here are based entirely on the fixed-base mode, without considering dynamic coupling effects between a floating platform and the manipulator.

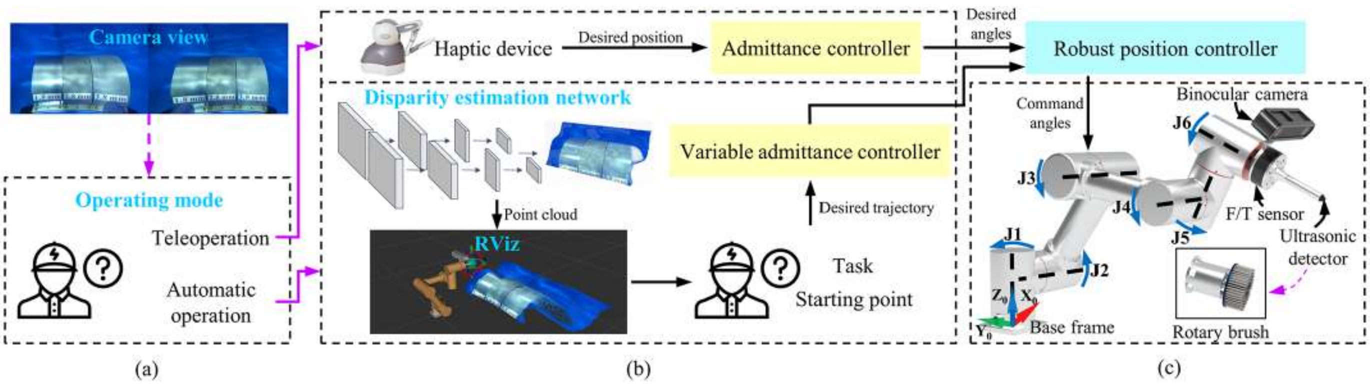

Fig. 1. System design and operation process: (a) selection of operating mode based on camera views, (b) composition of two operating modes developed using ROS2, (c) mechanical structure of the underwater manipulator.

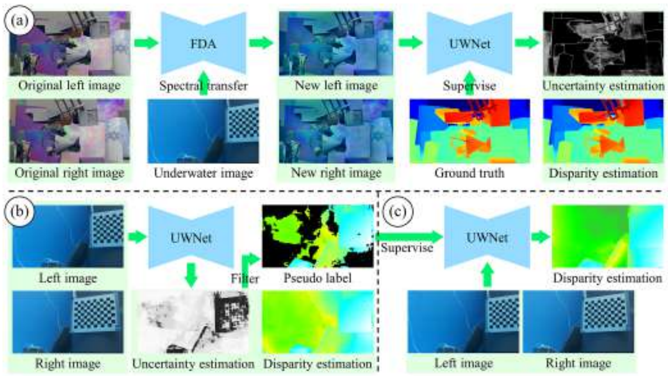

Fig. 2. Overall pre-training and adaptation framework of the UWNet, which consists of 3 steps: (a) pre-training the UWNet on the source domain, (b) pseudo-label generation on the target domain, and (c) domain adaptation with generated pseudo-labels.

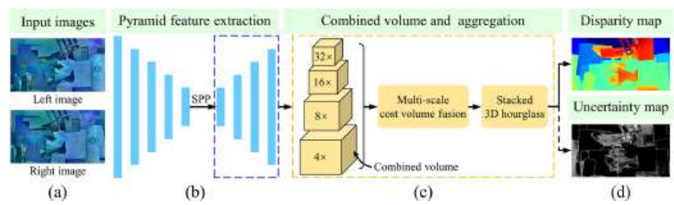

Fig. 3. Architecture of the proposed UWNet.

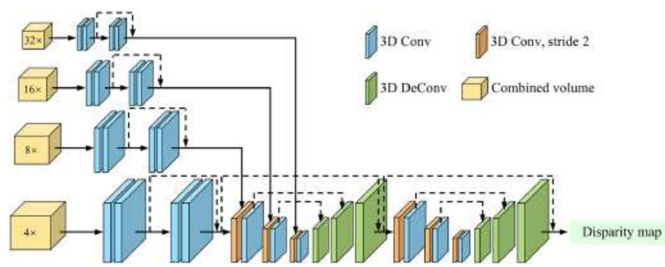

Fig. 4. Architecture of the cost volume fusion module and 3D hourglass network. Four different scale combined volumes ($i \in (2, 3, 4, 5)$) are fused to generate the disparity map.

III. Underwater Disparity Estimation

A. Framework Overview

Disparity maps provide pixel-wise depth information from stereo image pairs and are essential for perception tasks. However, underwater complex lighting and scattering effects reduce the reliability of disparity estimation. To address this, the proposed UWNet incorporates uncertainty map estimation to capture the confidence region of each predicted disparity and enable the generation of pseudo-labels for model adaptation. The entire UWNet pre-training and adaptation framework, as shown in Fig. 2, consists of the following three steps:

1) Pre-training the UWNet on source domain: First, Fourier Domain Adaptation (FDA) [30], a domain adaptation technique for aligning feature distributions between source and target domains, is applied to adapt the image pairs in the SceneFlow dataset [31] (i.e., the source domain) to the style of underwater images (i.e., the target domain). Leveraging these newly transformed image pairs and their corresponding ground truth disparities, the UWNet is pre-trained to simultaneously estimate disparities and uncertainties. This step establishes a baseline model that has learned general stereo matching features and serves as a foundation for further adaptation.

2) Pseudo-label generation on target domain: After pre-training UWNet, an uncertainty-based pseudo-label generation strategy is employed for domain adaptation, as detailed in Section III-B3. Stereo image pairs of underwater structure are input into the pre-trained UWNet to generate disparity and uncertainty maps. This uncertainty estimation is used to filter out high-uncertainty regions in the disparity map, yielding a sparse yet reliable pseudo-label that preserves only the most confident points and mitigate the adverse effects of underwater scenes.

3) Domain adaptation with generated Pseudo-label: The pseudo-labels generated in the previous step are then used to retrain UWNet, allowing it to adapt to underwater environments and improve disparity estimation accuracy. By fine-tuning UWNet on the underwater structure dataset with these pseudo-labels, its generalization capabilities are enhanced, enabling more precise underwater disparity estimation.

B. Uncertainty-based Underwater Disparity Estimation Network (UWNet)

To achieve robust disparity estimation in underwater environments, UWNet is developed based on GwcNet, with several key enhancements: replacing the feature extraction backbone with a UNet-like network, strengthening the combined volume module, and incorporating the uncertainty estimation module from UCFNet (see Fig. 3).

1) Pyramid feature extraction: Stereo image pairs are first processed by a pyramid feature extraction module built upon a UNet-like encoder-decoder architecture [19] (see Fig. 3(a) and (b)). Unlike the ResNet-based backbone of GwcNet, this design integrates multi-scale residual blocks with spatial pyramid pooling (SPP) [32] to enhance hierarchical context learning, while skip connections combine low-level and high-level features to preserve fine-grained details. Feature maps are extracted at scales of $1/2^i$ (for $i = 2$ to 5), corresponding to 1/4, 1/8, 1/16, and 1/32 of the input resolution.

2) Combined volume and aggregation: Unlike GwcNet, which constructs a single 1/4-scale combined volume, UWNet builds four multi-scale combined volumes to enhance robustness and generalization (see Fig. 3(c)). This design fuses high-resolution features for fine-grained details and low-resolution features for contextual information. At each scale, the combined cost volume (V^i_{comb}) is formed by concatenating feature concatenation (V^i_{concat}) and group-wise correlation (V^i_{gwc}):

$$\begin{aligned} V^i_{concat}(d^i, x, y, f) &= f^i_L(x, y) || f^i_R(x - d^i, y) \\ V^i_{gwc}(d^i, x, y, g) &= \frac{1}{N^i_c/N_g} \left\langle f^{ig}_l(x, y), f^{ig}_r(x - d^i, y) \right\rangle \\ V^i_{comb} &= V^i_{concat} || V^i_{gwc} \end{aligned} \tag{1}$$

where $||$ denotes concatenation, f^i_L and f^i_R are the left and right feature maps at scale i, d^i is the disparity searching index, and x, y are the spatial coordinates. N^i_c represents the channels of extracted features, N_g is the number of groups, and $\langle , \rangle$ indicates the inner product.

Subsequently, a multi-scale cost volume fusion method integrates the combined volumes across scales, followed by a 3D hourglass network for further aggregation, as shown in Fig. 4. In this process, each cost volume is first regularized by 3D convolutions, and higher-resolution volumes are progressively downsampled and concatenated with lower-resolution ones until a 1/32-scale representation is obtained. Finally, two 3D transposed convolutions and a 3D hourglass network are applied to upsample and aggregate the fused cost volume.

3) Uncertainty estimation for pseudo-label generation: Both the disparity and uncertainty map are predicted from the fused cost volume (see Fig. 3(d)). The uncertainty estimation is utilized to generate pseudo-labels, which serve as supervision for adapting the pre-trained model M_{pre}, initially trained on the source domain as detailed in Section III-A1, to the target domain. Specifically, M_{pre} takes target domain image pairs (I_L, I_R) as input and outputs a dense disparity map D_{pre} and an uncertainty map U. Unreliable points in D_{pre} are filtered out based on the uncertainty threshold, yielding a sparse yet reliable disparity maps D_f as pseudo-labels:

$$D_{pre} = M_{pre}(I_L, I_R), D_f = \{d \in D_{pre} : U < \mu + n\sigma\} \tag{2}$$

where $n = 0.3$ is the threshold coefficient for filtered disparity map D_f, μ and σ represent the mean and standard deviation of the uncertainty for each image. The configuration allows the threshold to adaptively adjust, preventing the filtered disparity map from having excessively many or too few valid values.

4) Loss function: The cross-entropy loss is adopted for training the uncertainty estimation module. The uncertainty loss L_u is defined based on the ground truth uncertainty U_{gt} and the predicted uncertainty U:

$$\begin{aligned} L_u &= U_{gt} \log U + (1 - U_{gt}) \log(1 - U) \\ U_{gt} &= \begin{cases} 1, & if \left| D_{gt} - \hat{D} \right| > 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{3}$$

where $\hat{D}$ denotes the predicted disparity and D_{gt} is the ground truth disparity. Here, $U_{gt} = 1$ marks pixels where the predicted disparity significantly deviates from the ground truth, indicating high uncertainty. For disparity estimation, a smooth L_1 loss is employed, consistent with that used in GwcNet.

IV. UNDERWATER ROBUST CONTROL

The robust controller for the underwater manipulator comprises an inner-loop position controller for precise trajectory tracking and an outer-loop force controller for compliant interaction and accurate force regulation, as shown in Fig. 5.

A. Inner-Loop Robust Position Controller Based on USDE and ST SMC

Compared with robots operating on land, underwater manipulators must contend with both intrinsic system uncertainties and hydrodynamic disturbances. Accurately modeling these effects is difficult, leading to complex control derivation and process [28]. Therefore, all hydrodynamic effects and model uncertainties can be treated as a single external disturbance $\boldsymbol{\tau}_d \in \mathbb{R}^n$. External measured forces are denoted by the Cartesian space wrenches $\hat{\boldsymbol{F}}_e = [\boldsymbol{f}^T_e, \boldsymbol{m}^T_e]^T \in \mathbb{R}^6$ consisting of translational forces $\boldsymbol{f}_e \in \mathbb{R}^3$ and rotational torques $\boldsymbol{m}_e \in \mathbb{R}^3$. The dynamic model of the underwater manipulator, which accounts for $\hat{\boldsymbol{F}}_e$ is given by:

$$\boldsymbol{M}(\boldsymbol{q})\ddot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{G}(\boldsymbol{q}) = \boldsymbol{\tau} + \boldsymbol{\tau}_d - \boldsymbol{J}^\top(\boldsymbol{q})\hat{\boldsymbol{F}}_e \tag{4}$$

where $\boldsymbol{q} \in \mathbb{R}^n$ are the joint positions, $\dot{\boldsymbol{q}} \in \mathbb{R}^n$ are the joint velocities, $\ddot{\boldsymbol{q}} \in \mathbb{R}^n$ are the joint accelerations, $\boldsymbol{M} \in \mathbb{R}^{n \times n}$ is the inertia matrix, $\boldsymbol{C} \in \mathbb{R}^{n \times n}$ is the Coriolis and centrifugal matrix, $\boldsymbol{G} \in \mathbb{R}^n$ is the gravity item, $\boldsymbol{\tau} \in \mathbb{R}^n$ is the control torque, and $\boldsymbol{J}^\top \in \mathbb{R}^{n \times 6}$ denotes Jacobian matrix.

The USDE is modified to estimate disturbances affected by contact forces. First, the following auxiliary variables are defined to reformulate (4):

$$\begin{cases} \boldsymbol{N}(\boldsymbol{q}, \dot{\boldsymbol{q}}) = \boldsymbol{M}(\boldsymbol{q})\dot{\boldsymbol{q}} \\ \boldsymbol{W}(\boldsymbol{q}, \dot{\boldsymbol{q}}) = -\dot{\boldsymbol{M}}(\boldsymbol{q})\dot{\boldsymbol{q}} + \boldsymbol{C}(\boldsymbol{q}, \dot{\boldsymbol{q}})\dot{\boldsymbol{q}} + \boldsymbol{G}(\boldsymbol{q}) \end{cases} . \tag{5}$$

Then, (4) is rewritten as

$$\dot{\boldsymbol{N}} + \boldsymbol{W} = \boldsymbol{\tau} + \boldsymbol{\tau}_d - \boldsymbol{J}^\top \hat{\boldsymbol{F}}_e. \tag{6}$$

Since $\boldsymbol{q}, \dot{\boldsymbol{q}}$ are measurable, $\boldsymbol{N}, \boldsymbol{W}$ are accessible. However, the joint acceleration $\ddot{\boldsymbol{q}}$ in $\dot{\boldsymbol{N}}$ is difficult to measure, while the time derivative of $\dot{\boldsymbol{q}}$ is sensitive to sensor noise. To effectively handle measurement noise and avoid direct differentiation, a

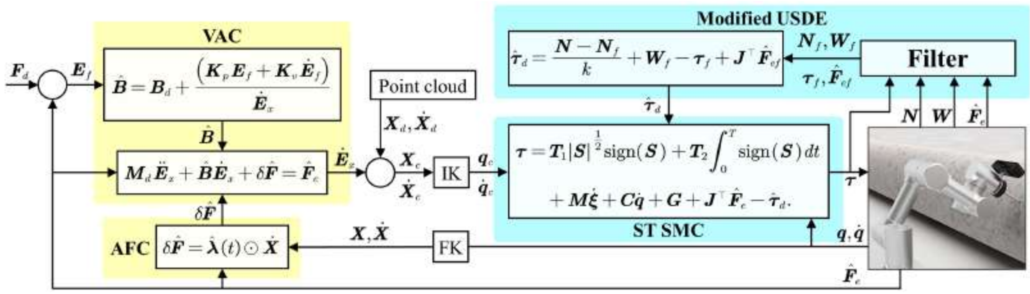

Fig. 5. Overview of the proposed controller, including an inner-loop robust position controller (enhanced with modified USDE and ST SMC) and an outer-loop variable admittance controller (VAC) with adaptive feedback compensation (AFC). IK means inverse kinematics, and FK means forward kinematics.

series of first-order low-pass filters [23], [24] are employed to obtain smooth estimates:

$$\begin{cases} k\dot{\boldsymbol{N}}_f + \boldsymbol{N}_f = \boldsymbol{N}, & \boldsymbol{N}_f|_{t=0} = \boldsymbol{0} \\ k\dot{\boldsymbol{W}}_f + \boldsymbol{W}_f = \boldsymbol{W}, & \boldsymbol{W}_f|_{t=0} = \boldsymbol{0} \\ k\dot{\boldsymbol{\tau}}_f + \boldsymbol{\tau}_f = \boldsymbol{\tau}, & \boldsymbol{\tau}_f|_{t=0} = \boldsymbol{0} \\ k\dot{\hat{\boldsymbol{F}}}_{ef} + \hat{\boldsymbol{F}}_{ef} = \hat{\boldsymbol{F}}_e, & \hat{\boldsymbol{F}}_{ef}|_{t=0} = \boldsymbol{0} \end{cases} \tag{7}$$

where $k > 0$ is a scalar filter parameter, $\boldsymbol{N}_f$, $\boldsymbol{W}_f$, $\boldsymbol{\tau}_f$, $\hat{\boldsymbol{F}}_{ef}$ are the filtered values of $\boldsymbol{N}$, $\boldsymbol{W}$, $\boldsymbol{\tau}$, $\hat{\boldsymbol{F}}_e$, respectively.

By substituting the filtered values from (7) into the reformulated model (6), the modified USDE is designed as:

$$\hat{\boldsymbol{\tau}}_d = \frac{\boldsymbol{N} - \boldsymbol{N}_f}{k} + \boldsymbol{W}_f - \boldsymbol{\tau}_f + \boldsymbol{J}^\top \hat{\boldsymbol{F}}_{ef}. \tag{8}$$

$\boldsymbol{N}$ and $\boldsymbol{W}$ are variables related to $\boldsymbol{q}$ and $\dot{\boldsymbol{q}}$, where the measured $\dot{\boldsymbol{q}}$ from joint encoders inevitably contains noise signals. These signals will be amplified if the filter coefficient k is too small. Hence, the method of reconstructing $\dot{\boldsymbol{q}}$ from the enhanced USDE [24] is employed, resulting in a new variable, $\dot{\boldsymbol{q}}_f$, which replaces $\dot{\boldsymbol{q}}$ in (8). This process avoids measuring joint accelerations and eliminates the need to compute the inverse of the inertia matrix, thereby enhancing computational efficiency. The estimation error of the disturbance is given by:

$$\begin{aligned} \tilde{\boldsymbol{\tau}}_d =& \boldsymbol{\tau}_d - \hat{\boldsymbol{\tau}}_d = k\left(\frac{\dot{\boldsymbol{N}} - \dot{\boldsymbol{N}}_f}{k} + \dot{\boldsymbol{W}}_f - \dot{\boldsymbol{\tau}}_f + \boldsymbol{J}^\top \dot{\hat{\boldsymbol{F}}}_{ef}\right) \\ =& k\dot{\hat{\boldsymbol{\tau}}}_d \end{aligned} \tag{9}$$

which indicates that the estimated disturbance $\hat{\boldsymbol{\tau}}_d$ is a low-pass filtered version of the actual disturbance $\boldsymbol{\tau}_d$. To investigate the convergence of the disturbance estimation error $\tilde{\boldsymbol{\tau}}_d$, the following Lyapunov function is defined:

$$V_d = \frac{1}{2}\tilde{\boldsymbol{\tau}}_d^\top \tilde{\boldsymbol{\tau}}_d. \tag{10}$$

Calculating its differential with respect to time and applying Young's inequality on $\tilde{\boldsymbol{\tau}}_d^\top \dot{\boldsymbol{\tau}}_d$, the inequality is obtained:

$$\begin{aligned} \dot{V}_d =& -\frac{1}{k}\tilde{\boldsymbol{\tau}}_d^\top \tilde{\boldsymbol{\tau}}_d + \tilde{\boldsymbol{\tau}}_d^\top \dot{\boldsymbol{\tau}}_d \le -\frac{1}{k}\tilde{\boldsymbol{\tau}}_d^\top \tilde{\boldsymbol{\tau}}_d + \frac{1}{2k}\tilde{\boldsymbol{\tau}}_d^\top \tilde{\boldsymbol{\tau}}_d + \frac{k}{2}\tau_{d_0}^2 \\ \le& -\frac{1}{k}V_d + \frac{k}{2}\tau_{d_0}^2 \end{aligned} \tag{11}$$

where $\tau_{d_0} > 0$ is the constant bound of the actual disturbance derivative $\dot{\boldsymbol{\tau}}_d$, i.e., $\sup_{t\ge 0}\|\dot{\boldsymbol{\tau}}_d\| \le \tau_{d_0}$, V_d is proven to be bounded and convergent. Equation (11) can yield that $||\tilde{\boldsymbol{\tau}}_d|| \le \sqrt{||\tilde{\boldsymbol{\tau}}_d|_{t=0}||^2 e^{-t/k} + k^2\tau_{d_0}^2}$, which indicates that $\tilde{\boldsymbol{\tau}}_d$ will exponentially converge to a bound around zero and the estimation $\hat{\boldsymbol{\tau}}_d \to \boldsymbol{\tau}_d$ holds for $k \to 0$ and/or $\tau_{d_0} \to 0$.

By utilizing the disturbance estimation $\hat{\boldsymbol{\tau}}_d$, the control torque $\boldsymbol{\tau}$ is obtained from (4) as follows:

$$\boldsymbol{\tau} = \boldsymbol{M}\boldsymbol{u}_p + \boldsymbol{C}\dot{\boldsymbol{q}} + \boldsymbol{G} + \boldsymbol{J}^\top \hat{\boldsymbol{F}}_e - \hat{\boldsymbol{\tau}}_d \tag{12}$$

where $\boldsymbol{u}_p$ is the control law based on the current and desired states to track the desired trajectory. To design a robust control law based on ST SMC, a variable $\boldsymbol{S} \in \mathbb{R}^n$ is defined as

$$\boldsymbol{S} = \dot{\boldsymbol{e}} + \boldsymbol{\eta}\boldsymbol{e} \tag{13}$$

where $\boldsymbol{\eta} \in \mathbb{R}^{n\times n}$ is a positive diagonal matrix of control coefficient, $\boldsymbol{e} = \boldsymbol{q}_c - \boldsymbol{q}$ represents the tracking error, and $\boldsymbol{q}_c$ denotes the desired command. The variable $\boldsymbol{S}$ serves as the sliding manifold in the robust controller, with the reaching law of the ST algorithm

$$\dot{\boldsymbol{S}} = -\boldsymbol{T}_1|\boldsymbol{S}|^{\frac{1}{2}}\mathrm{sign}(\boldsymbol{S}) - \boldsymbol{T}_2\int_0^T \mathrm{sign}(\boldsymbol{S})dt \tag{14}$$

where $\boldsymbol{T}_1, \boldsymbol{T}_2 \in \mathbb{R}^{n\times n}$ are diagonal matrices of positive control gains, T denotes the control cycle duration. From (13), it follows that:

$$\boldsymbol{M}\dot{\boldsymbol{S}} = \boldsymbol{M}(\ddot{\boldsymbol{q}}_c - \ddot{\boldsymbol{q}} + \boldsymbol{\eta}\dot{\boldsymbol{e}}) = \boldsymbol{M}(\dot{\boldsymbol{\xi}} - \ddot{\boldsymbol{q}}). \tag{15}$$

Combining (12), (14), (15), the robust control law is

$$\boldsymbol{u}_p = \dot{\boldsymbol{\xi}} + \boldsymbol{T}_1|\boldsymbol{S}|^{\frac{1}{2}}\mathrm{sign}(\boldsymbol{S}) + \boldsymbol{T}_2\int_0^T \mathrm{sign}(\boldsymbol{S})dt. \tag{16}$$

Finally, the control torque is computed as:

$$\begin{aligned} \boldsymbol{\tau} =& \boldsymbol{T}_1|\boldsymbol{S}|^{\frac{1}{2}}\mathrm{sign}(\boldsymbol{S}) + \boldsymbol{T}_2\int_0^T \mathrm{sign}(\boldsymbol{S})dt + \boldsymbol{M}\dot{\boldsymbol{\xi}} \\ &+ \boldsymbol{C}\dot{\boldsymbol{q}} + \boldsymbol{G} + \boldsymbol{J}^\top \hat{\boldsymbol{F}}_e - \hat{\boldsymbol{\tau}}_d. \end{aligned} \tag{17}$$

In this configuration, the modified USDE compensates for most system uncertainties and external disturbances, while the ST SMC corrects control torque deviations due to errors.

B. Outer-Loop Variable Admittance Controller with Adaptive Feedback Compensation

The equation for traditional admittance control is

$$\boldsymbol{M}_d\ddot{\boldsymbol{E}}_x + \boldsymbol{B}_d\dot{\boldsymbol{E}}_x + \boldsymbol{K}_d\boldsymbol{E}_x = -\boldsymbol{E}_f, \tag{18}$$

where $\boldsymbol{M}_d$, $\boldsymbol{B}_d$, and $\boldsymbol{K}_d \in \mathbb{R}^{n\times n}$ represent the desired mass, damping, and stiffness matrices, respectively, all of which are symmetric and positive definite. $\boldsymbol{E}_x = \boldsymbol{X}_d - \boldsymbol{X}_c \in \mathbb{R}^n$ is the position correction. $\boldsymbol{E}_f = \boldsymbol{F}_d - \hat{\boldsymbol{F}}_e \in \mathbb{R}^n$ where $\boldsymbol{F}_d$ denotes the desired contact force. The compliant Cartesian position reference $\boldsymbol{X}_c$ from the out-loop admittance controller is converted to joint position reference $\boldsymbol{q}_c$ via inverse kinematics and tracked by the inner-loop position controller. In traditional admittance control, contact forces are indirectly regulated by specifying desired impedance dynamics, making accurate force tracking challenging due to uncertainties in environmental parameters such as contact location and stiffness.

For compliant interaction and force tracking with minimal overshoot and oscillation, the variable admittance controller with adaptive feedback compensation is designed as:

$$\underbrace{\boldsymbol{M}_d\ddot{\boldsymbol{E}}_x + \hat{\boldsymbol{B}}\dot{\boldsymbol{E}}_x}_{variable\ admittance} + \underbrace{\delta\hat{\boldsymbol{F}}}_{adaptive} = \hat{\boldsymbol{F}}_e \tag{19}$$

where $\hat{\boldsymbol{B}}$ is a variable damping matrix that adapts based on the force error, $\delta\hat{\boldsymbol{F}}$ is the adaptive item. $\hat{\boldsymbol{B}}$ is defined as:

$$\hat{\boldsymbol{B}} = \boldsymbol{B}_d + \frac{\left(\boldsymbol{K}_p\boldsymbol{E}_f + \boldsymbol{K}_v\dot{\boldsymbol{E}}_f\right)}{\dot{\boldsymbol{E}}_x} \tag{20}$$

where $\boldsymbol{K}_p$ and $\boldsymbol{K}_v$ are diagonal proportional-derivative (PD) gains. Unlike traditional admittance control, this approach omits stiffness effects and adaptively adjusts damping, refining the inner-loop inputs to achieve effective force tracking.

The purpose of the adaptive feedback compensation is to mitigate impact during the transition phases and regulate the uncertain forces between the robot and environment. $\delta\hat{\boldsymbol{F}}$ is defined as a function of robot velocity and contact force:

$$\delta\hat{\boldsymbol{F}} = \hat{\boldsymbol{\lambda}}(t) \odot \dot{\boldsymbol{X}} \tag{21}$$

where $\hat{\boldsymbol{\lambda}}(t)$ is the adaptive control gain, adjusted online based on the magnitude of the contact force:

$$\hat{\boldsymbol{\lambda}}(t) = \beta|\hat{\boldsymbol{F}}_e| \geq 0 \tag{22}$$

with $\beta = 0.02$ as a positive constant. Before contact ($|\hat{\boldsymbol{F}}_e| = 0$), $\delta\hat{\boldsymbol{F}}$ remains zero regardless of velocity. After contact ($|\hat{\boldsymbol{F}}_e| \neq 0$), the adaptive feedback compensator addresses two distinct scenarios. First, during the transition from non-contact to contact, characterized by high contact force and robot velocity, $\delta\hat{\boldsymbol{F}}$ contributes to robot stabilization by reducing the impact and minimizing force overshoot. Second, when the end-effector maintains continuous contact at typically low velocities, $\delta\hat{\boldsymbol{F}}$ diminishes, allowing the variable admittance controller to dominate and ensure precise force tracking.

C. Stability Analysis

The stability and convergence of the inner-loop controller are first analyzed, followed by the stability analysis of the outer-loop and the complete closed-loop system based on the inner-loop results. For the robust position controller, substituting (17) into (4) yields:

$$\boldsymbol{M}\dot{\boldsymbol{S}} = -\boldsymbol{M}\boldsymbol{T}_1|\boldsymbol{S}|^{\frac{1}{2}}\mathrm{sign}(\boldsymbol{S}) + \boldsymbol{M}\boldsymbol{\Omega} + \boldsymbol{M}\boldsymbol{\zeta}_1 + \boldsymbol{M}\int_0^T \boldsymbol{\zeta}_2 dt \tag{23}$$

with

$$\dot{\boldsymbol{\Omega}} = -\boldsymbol{T}_2\mathrm{sign}(\boldsymbol{S}), \boldsymbol{M}\boldsymbol{\zeta}_1 = (\boldsymbol{M} - \boldsymbol{I})\boldsymbol{T}_1|\boldsymbol{S}|^{\frac{1}{2}}\mathrm{sign}(\boldsymbol{S}),$$
$$\boldsymbol{M}\int_0^T \boldsymbol{\zeta}_2 dt = (\boldsymbol{I} - \boldsymbol{M})\boldsymbol{\Omega} - \tilde{\boldsymbol{\tau}}_d \tag{24}$$

where $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{R}^n$ are auxiliary terms. Because $\boldsymbol{M}$ and $\tilde{\boldsymbol{\tau}}_d$ are bounded, there exist $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2$ with properties of $\sup_{t\geq 0}|\zeta_{1_i}| \leq \phi_{1_i}|S_i|^{\frac{1}{2}}$ and $\sup_{t\geq 0}|\zeta_{2_i}| \leq \phi_{2_i}$ for some constants $\phi_{1_i}, \phi_{2_i} > 0$. Thus, (23) is reformulated as:

$$\begin{cases} \dot{\boldsymbol{S}} = -\boldsymbol{T}_1|\boldsymbol{S}|^{\frac{1}{2}}\mathrm{sign}(\boldsymbol{S}) + \boldsymbol{\Omega\prime} + \boldsymbol{\zeta}_1 \\ \dot{\boldsymbol{\Omega}}\prime = -\boldsymbol{T}_2\mathrm{sign}(\boldsymbol{S}) + \boldsymbol{\zeta}_2 \end{cases}. \tag{25}$$

Then, considering a Lyapunov function candidate as

$$V_P = \sum_{i=1}^{n} V_{S_i} + \frac{1}{2}\tilde{\boldsymbol{\tau}}_d^\top\tilde{\boldsymbol{\tau}}_d = \sum_{i=1}^{n} X_i^\top P_i X_i + \frac{1}{2}\tilde{\boldsymbol{\tau}}_d^\top\tilde{\boldsymbol{\tau}}_d \tag{26}$$
$$X_i = \begin{bmatrix} |S_i|^{\frac{1}{2}}\mathrm{sign}(S_i) \\ \Omega_i' \end{bmatrix}, P_i = \frac{1}{2}\begin{bmatrix} 4T_{2_i} + T_{1_i}^2 & -T_{1_i} \\ -T_{1_i} & 2 \end{bmatrix}$$

where X_i is an auxiliary variable, P_i is a constant symmetric and positive definite matrix, T_{1_i}, T_{2_i} correspond to the diagonal element of $\boldsymbol{T}_1, \boldsymbol{T}_2$.

The element of $V_S = \sum_{i=1}^{n} V_{S_i}$ satisfies the inequality:

$$\lambda_{\min}(P_i)||X_i||_2^2 \leq V_{S_i} \leq \lambda_{\max}(P_i)||X_i||_2^2. \tag{27}$$

Thus,

$$|S_i|^{\frac{1}{2}} \leq ||X_i||_2 \leq \frac{V_{S_i}^{\frac{1}{2}}}{\lambda_{\min}^{\frac{1}{2}}(P_i)}, \frac{V_{S_i}}{\lambda_{\max}(P_i)} \leq ||X_i||_2^2. \tag{28}$$

Taking the time derivative of X_i and applying the (25) yields:

$$\dot{X}_i = -\frac{1}{2|S_i|^{\frac{1}{2}}}\begin{bmatrix} T_{1_i} & -1 \\ 2T_{2_i} & 0 \end{bmatrix} X_i + \frac{1}{2|S_i|^{\frac{1}{2}}}\begin{bmatrix} \zeta_{1_i} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \zeta_{2_i} \end{bmatrix}. \tag{29}$$

According to (28) and (29), the time derivative of V_{S_i} satisfies

$$\dot{V}_{S_i} = 2X_i^\top P_i\dot{X}_i \leq -\frac{1}{|S_i|^{\frac{1}{2}}}X_i^\top Q_i X_i$$
$$\leq -\frac{\lambda_{\min}^{\frac{1}{2}}(P_i)}{V_{S_i}^{\frac{1}{2}}}\lambda_{\min}(Q_i)\frac{V_{S_i}}{\lambda_{\max}(P_i)} = -\psi_i V_{S_i}^{\frac{1}{2}} \tag{30}$$

with

$$Q_i = \frac{T_{1_i}}{2}\begin{bmatrix} 2T_{2_i} + T_{1_i}^2 - \left(\frac{4T_{2_i}}{T_{1_i}} + T_{1_i}\right)\phi_{1_i} - 2\phi_{2_i} & * \\ -\left(T_{1_i} + 2\phi_{1_i} + 2\frac{\phi_{2_i}}{T_{1_i}}\right) & 1 \end{bmatrix}$$
$$\psi_i = \frac{\lambda_{\min}^{\frac{1}{2}}(P_i)\lambda_{\min}(Q_i)}{\lambda_{\max}(P_i)} \tag{31}$$

where Q_i is a symmetric and positive definite matrix by selecting appropriate gains T_{1_i}, T_{2_i}. Therefore, $\dot{V}_P$ is

$$\dot{V}_P = \sum_{i=1}^{n}\dot{V}_{S_i} + \tilde{\boldsymbol{\tau}}_d^\top\dot{\tilde{\boldsymbol{\tau}}}_d = \dot{V}_S + \tilde{\boldsymbol{\tau}}_d^\top\left(-\frac{1}{k}\tilde{\boldsymbol{\tau}}_d + \dot{\boldsymbol{\tau}}_d\right) \tag{32}$$
$$\leq \dot{V}_S - \frac{\sqrt{2}}{2k}\left(\frac{1}{2}\tilde{\boldsymbol{\tau}}_d^\top\tilde{\boldsymbol{\tau}}_d\right)^{\frac{1}{2}} + \frac{1}{8k} + \frac{k}{2}\tau_{d_0}^2 \leq -\alpha_p V_P^{\frac{1}{2}} + \beta_p$$

where $\alpha_p = \min\{\psi_i, \frac{\sqrt{2}}{2k}\}, \beta_p = \frac{1}{8k} + \frac{k}{2}\tau_{d_0}^2$, and both are positive. Thus, the position controller is stable. The tracking error $\boldsymbol{S}$ and disturbance estimation error $\tilde{\boldsymbol{\tau}}_d$ will converge to a small compact set around zero in finite time, with the proper choice of the control parameters $\boldsymbol{T}_1$, $\boldsymbol{T}_2$, and k. An increase in τ_{d_0} enlarges the bounded set which $\boldsymbol{S}$ converges to, leading to a larger tracking error of the inner-loop position controller.

Based on the inner-loop stability analysis, consider the following Lyapunov function candidate for the outer loop:

$$V_F = \frac{1}{2}\dot{\boldsymbol{E}}_x^\top \boldsymbol{M}_d \dot{\boldsymbol{E}}_x,$$
$$\dot{V}_F = \dot{\boldsymbol{E}}_x^\top \boldsymbol{M}_d \ddot{\boldsymbol{E}}_x = -\dot{\boldsymbol{E}}_x^\top \hat{\boldsymbol{B}} \dot{\boldsymbol{E}}_x - \dot{\boldsymbol{E}}_x^\top \delta \hat{\boldsymbol{F}} + \dot{\boldsymbol{E}}_x^\top \hat{\boldsymbol{F}}_e \ . \tag{33}$$

Applying the Cauchy-Schwarz and Young's inequalities (for any $\varepsilon, \rho > 0$), defining $\lambda_B = \lambda_{\min}(\hat{\boldsymbol{B}})$, and selecting the parameters ε, ρ such that $\alpha_f = \lambda_B - \frac{\varepsilon+\rho}{2} > 0$, we obtain

$$\dot{V}_F \leq -\alpha_f ||\dot{\boldsymbol{E}}_x||^2 + \frac{1}{2\varepsilon}||\hat{\boldsymbol{F}}_e||^2 + \frac{1}{2\rho}||\hat{\boldsymbol{\lambda}}(t) \odot \dot{\boldsymbol{X}}||^2 \tag{34}$$

where $\hat{\boldsymbol{\lambda}}(t) = \beta|\hat{\boldsymbol{F}}_e|$ is bounded above by $\hat{\lambda}_{\max}$, $\dot{\boldsymbol{X}} = \dot{\boldsymbol{X}}_c + (\dot{\boldsymbol{X}} - \dot{\boldsymbol{X}}_c), \dot{\boldsymbol{X}}_c = \dot{\boldsymbol{X}}_d - \dot{\boldsymbol{E}}_x$. Thus,

$$||\hat{\boldsymbol{\lambda}}(t) \odot \dot{\boldsymbol{X}}||^2 \leq 3\hat{\lambda}_{\max}^2 \left(||\dot{\boldsymbol{X}}_d||^2 + ||\dot{\boldsymbol{E}}_x||^2 + ||\dot{\boldsymbol{X}} - \dot{\boldsymbol{X}}_c||^2\right) \tag{35}$$

where $||\dot{\boldsymbol{X}} - \dot{\boldsymbol{X}}_c||$ denotes the inner-loop tracking error, and $\sup_{t\geq 0}||\dot{\boldsymbol{X}} - \dot{\boldsymbol{X}}_c|| \leq \varrho$, ϱ is a positive constant. Substituting (35) into (34) and choosing ρ sufficiently large such that $\alpha_{fx} = \alpha_f - \frac{3\hat{\lambda}_{\max}^2}{2\rho} > 0$:

$$\dot{V}_F \leq -\alpha_{fx}||\dot{\boldsymbol{E}}_x||^2 + \varSigma, \tag{36}$$

where $\varSigma = \frac{1}{2\varepsilon}||\hat{\boldsymbol{F}}_e||^2 + \frac{3\hat{\lambda}_{\max}^2}{2\rho}||\dot{\boldsymbol{X}}_d||^2 + \frac{3\hat{\lambda}_{\max}^2}{2\rho}\varrho^2$ is a constant upper bound determined jointly by $||\hat{\boldsymbol{F}}_e||^2, ||\dot{\boldsymbol{X}}_d||^2$ and ϱ^2. Therefore, the states $\boldsymbol{E}_x, \dot{\boldsymbol{E}}_x$ of the outer-loop controller are uniformly ultimately bounded. When τ_{d_0} increases, the inner-loop tracking error also increases, which enlarges the ultimate bound of $\boldsymbol{E}_x$, thereby indicating a cascaded amplification relationship from τ_{d_0} to $\boldsymbol{E}_x$.

Based on the above analysis, the entire closed-loop system can be regarded as a cascaded structure. The outer loop generates the reference $\boldsymbol{X}_c$, while the inner loop tracks the corresponding joint position $\boldsymbol{q}_c$. The residual error of the inner loop affects $\hat{\boldsymbol{F}}_e$ and $\hat{\boldsymbol{\lambda}}(t) \odot \dot{\boldsymbol{X}}$, which act as bounded inputs to the outer loop. Therefore, the Lyapunov function candidate for the entire closed-loop system is chosen as:

$$V_C = \frac{1}{2}\tilde{\boldsymbol{\tau}}_d^\top \tilde{\boldsymbol{\tau}}_d + \sum_{i=1}^{n} V_{S_i} + \omega V_F, \tag{37}$$

where $\omega > 0$ is a weighting factor. Thus,

$$\dot{V}_C \leq -\alpha_p V_P^{\frac{1}{2}} - \omega\alpha_{fx}||\dot{\boldsymbol{E}}_x||^2 + \varPhi\left(\tau_{d_0}, k, ||\hat{\boldsymbol{F}}_e||, \hat{\lambda}_{\max}\right) \tag{38}$$

where $\varPhi(\cdot)$ is a non-negative constant composed of bounded quantities. Therefore, the entire closed-loop system is stable. As τ_{d_0} increases, the radius of the closed-loop error bound enlarges, while appropriate selection of $k, \boldsymbol{T}_1, \boldsymbol{T}_2$ can reduce the error bound to the desired range.

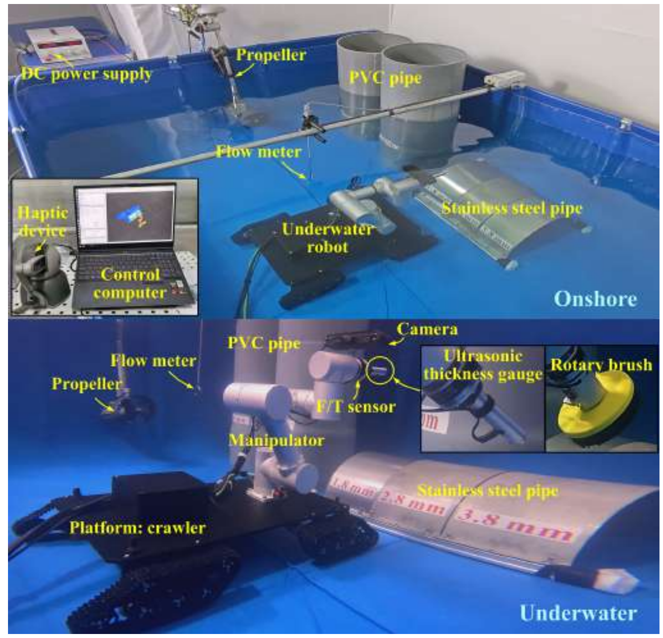

Fig. 6. Experimental setup for validating the performance of the designed manipulator system and methods in underwater inspection and maintenance.

Fig. 7. Example images of underwater dataset.

V. Experiments and Results

A. Experimental Setup and Protocol

The experiments were conducted in a 3m × 2.5m water tank, as shown in Fig. 6. The underwater manipulator, mounted on a crawler platform, is equipped with a waterproof ZED 2i stereo camera and a 6-axis F/T sensor (M3714BP, Sunrise Instruments) on the wrist. An ultrasonic thickness gauge and a rotary brush are used for thickness measurement and surface cleaning, respectively. All devices are connected to an onshore computer running Ubuntu 22.04 LTS and ROS2 Humble. The manipulator and F/T sensor operate at 500 Hz, and the camera streams at 30 fps. The test targets include 40 cm-diameter stainless steel pipes (1.8, 2.8, and 3.8 mm thick) and PVC pipes (6 and 9 mm thick). Different flow conditions are generated by a propeller and measured using a flow meter.

1) Underwater disparity estimation: UWNet and several state-of-the-art models are trained on the SceneFlow dataset, then evaluated on the KITTI [33] and FLSea [34] datasets. To further evaluate underwater disparity estimation in hydraulic structures, a dataset of 255 rectified underwater stereo image pairs was collected after underwater camera calibration to eliminate distortions and other underwater effects (see Fig. 7). Pseudo-labels generated from this dataset are used to fine-tune

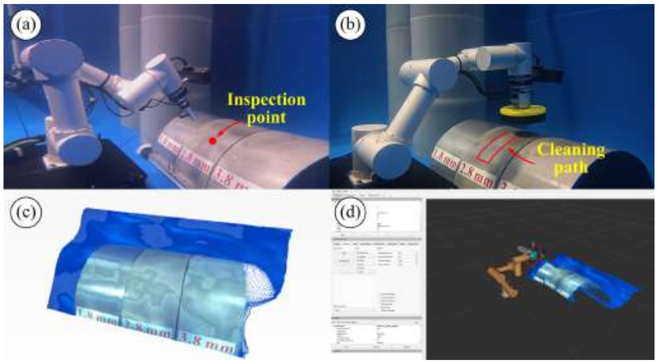

Fig. 8. Scenarios of (a) point inspection and (b) surface cleaning. (c) shows the point cloud of a stainless steel pipe, while (d) depicts the RViz interface.

the UWNet pre-trained on the SceneFlow dataset for domain adaptation. To effectively quantify depth estimation accuracy, the operator randomly selects 20 points from the pipe point cloud, generated by UWNet, as desired inspection positions for the manipulator. The point inspection scenario, the generated point cloud, and its display in RViz are shown in Fig. 8(a), (c), and (d). The actual inspection position is derived via forward kinematics at the moment the end-effector contacts the pipe.

2) Trajectory tracking: The trajectory tracking performance of the inner-loop position controller (eq.(17), denoted as USDE-ST) is validated under various water flow velocities. USDE-ST is configured with a filter coefficient $k = 0.05$ and the control gains: $\boldsymbol{\eta} = \mathrm{diag}([8\ 8\ 8\ 8\ 8\ 8])$, $\boldsymbol{T}_1 = \mathrm{diag}([2\ 2\ 2\ 1\ 1\ 1])$, and $\boldsymbol{T}_2 = \mathrm{diag}([6\ 6\ 6\ 2\ 2\ 2])$. The root-mean-square error (RMSE) of Cartesian position is introduced as a metric to compare the performance of different methods.

3) Force control, Inspection, and Maintenance: The outer-loop force controller (eq.(19)) is evaluated through pipe thickness inspection and surface cleaning experiments. In the inspection task (see Fig. 8(a)), operator selects points on the point cloud, and the manipulator performs ten ultrasonic measurements per pipe, using sound velocities of 5740 m/s for stainless steel and 2388 m/s for PVC. A micrometer measures the actual thickness as a reference. In the cleaning task (see Fig. 8(b)), RANSAC is used to filter point cloud noise and fit the cylindrical surface, after which a U-shaped cleaning path is executed from a selected starting point. The admittance parameters are set as $\boldsymbol{M}_d = \mathrm{diag}([10\ 10\ 10\ 1\ 1\ 1])$, $\boldsymbol{B}_d = \mathrm{diag}([400\ 400\ 400\ 40\ 40\ 40])$, and $\boldsymbol{K}_d = \mathrm{diag}([1000\ 1000\ 1000\ 100\ 100\ 100])$, with variable damping gains $\boldsymbol{K}_p = 120\boldsymbol{I}$ and $\boldsymbol{K}_v = 10\boldsymbol{I}$. A desired contact force of 10 N is applied along the z-axis, while traditional admittance control is used for the remaining axes.

B. Results

Table I compares the estimation accuracy of different models on the KITTI and FLSea datasets. The pre-trained UWNet outperforms GwcNet baseline, achieving relative improvements of 61.7% on KITTI2012, 54.1% on KITTI2015, and 57.4% on FLSea, demonstrating its stronger cross-domain generalization capability. However, its performance remains inferior to FoundationStereo. After domain adaptation using pseudo-labels generated from KITTI or FLSea, UWNet_adapt achieves further accuracy improvements and surpasses FoundationStereo on both datasets, confirming the effectiveness of the proposed domain adaptation framework across onshore and underwater scenarios. The results of disparity estimation and point cloud generation for underwater structures are depicted in Fig. 9. Compared to GwcNet and UWNet_pretrain, UWNet_adapt generates disparity maps with fewer artifacts and smoother transitions, particularly around object boundaries, while its point cloud exhibits sharper contours and more complete surface reconstructions for the slab and pipe.

When UWNet is applied to depth estimation of the pipe, Fig. 10 presents the errors between the actual positions and the corresponding point cloud positions, with the x-axis corresponding to the depth direction measured by the stereo camera. The depth errors span a wide range, approximately from -8 mm to 14 mm, and are widely scattered, whereas the errors along the y and z directions are constrained within a smaller range. Underwater disparity estimation experiments are conducted under clear water conditions and have yet to be validated across diverse aquatic environments, such as those with varying visibility and turbidity. Integrating advanced

TABLE I
ESTIMATION ACCURACY COMPARISON OF DIFFERENT MODELS TRAINED ON SCENEFLOW DATASET. IMP. DENOTES ACCURACY IMPROVEMENT.

Method	KITTI2012		KITTI2015		FLSea	
	D1_all (%)	Imp.	D1_all (%)	Imp.	RMSE (m)	Imp.
GwcNet [17]	12.0	-	12.2	-	2.532	-
CFNet [18]	4.7	60.8%	5.8	52.5%	1.203	52.5%
NMRF [35]	4.2	65.0%	5.1	58.2%	1.145	54.8%
FoundationStereo [36]	3.2	73.3%	4.9	59.8%	0.851	66.4%
UWNet_pretrain	4.6	61.7%	5.6	54.1%	1.078	57.4%
UWNet_adapt	3.2	73.3%	3.4	72.1%	0.694	72.6%

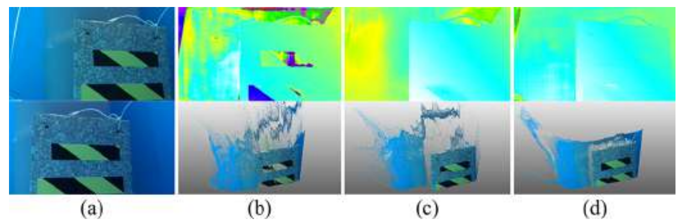

Fig. 9. Disparity estimation and point clouds of slab and pipe using different networks. (a) Stereo image pairs, (b), (c), (d) are predictions of GwcNet, UWNet_pretrain, and UWNet_adapt, respectively.

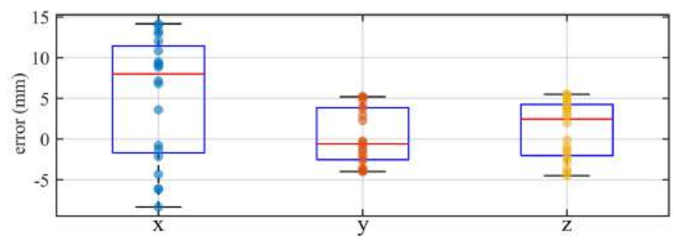

Fig. 10. Errors in estimating pipe surface point positions using UWNet.

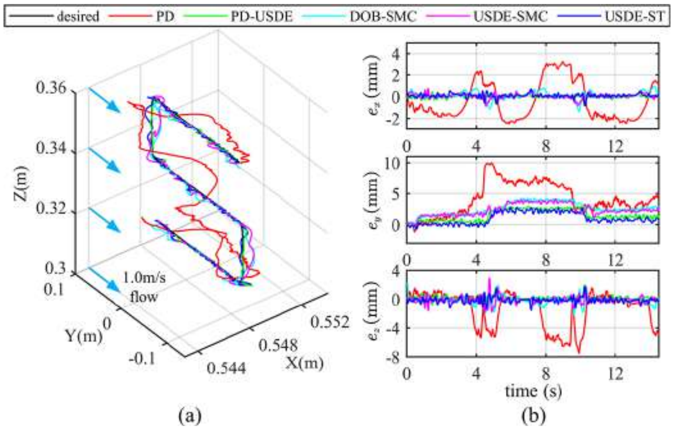

Fig. 11. Comparison of trajectory tracking performance between five different methods at 1.0 m/s flow: (a) Cartesian trajectories and (b) position errors.

TABLE II
STATISTICAL ERROR ANALYSIS OF POSITION CONTROLLERS AT DIFFERENT FLOW VELOCITIES (AVERAGE OF 5 TRIALS)

		PD	PD-USDE	DOB-SMC	USDE-SMC	USDE-ST
0.0m/s	RMSE (mm)	4.65	1.36	2.15	1.93	1.10
	Improvement	0	70.8%	53.8%	58.5%	76.3%
0.5m/s	RMSE (mm)	5.02	1.50	2.36	2.13	1.16
	Improvement	0	70.1%	53.0%	57.6%	76.9%
1.0m/s	RMSE (mm)	5.98	1.81	2.82	2.57	1.44
	Improvement	0	69.7%	52.8%	57.0%	75.9%

TABLE III
STATISTICAL ANALYSIS OF CONTACT FORCES FOR DIFFERENT ADMITTANCE CONTROLLERS AT 0.0 M/S FLOW (AVERAGE OF 5 TRIALS)

Method	Peak contact force (N)	Steady state		
		Settling time (s)	Fluctuation range (N)	Mean error (N)
VAC	-34.05	9.02	[-11.20, -7.49]	-1.06
MB-VAC	-26.27	10.89	[-11.25, -8.97]	-0.75
VAC+AFC	-23.90	4.16	[-11.07, -8.56]	-0.63

image enhancement methods to improve image contrast and clarity could enhance the reliability and robustness of UWNet.

In the trajectory tracking experiments, five methods are evaluated under varying flow conditions. Fig. 11 presents the results at 1.0 m/s flow. Due to the lack of compensation for model uncertainties and hydrodynamic disturbances, the PD controller shows large deviations from the desired path. In contrast, methods incorporating USDE markedly improve trajectory tracking accuracy. Moreover, the integration of ST SMC further reduces the position errors and enhances system robustness. Traditional SMC-based methods, such as USDE-SMC and DOB-SMC [22], exhibit lower tracking accuracy and more pronounced position fluctuations compared to USDE-ST. Table II shows that while DOB-SMC and USDE-SMC reduce RMSE, their improvements are limited. In contrast, USDE-ST substantially enhances trajectory tracking accuracy, achieving an RMSE of 1.44 mm at 1.0 m/s water flow and a 75.7% improvement over the baseline.

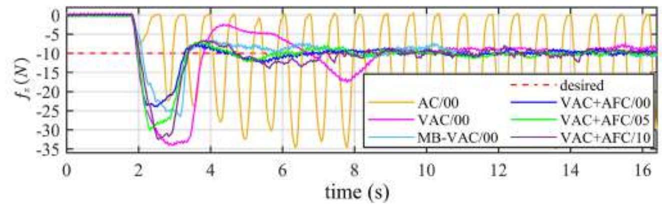

Fig. 12. Comparison of contact forces f_z during point inspection. AC means traditional admittance controller; /00, /05, and /10 correspond to flow speeds of 0.0 m/s, 0.5 m/s, and 1.0 m/s, respectively.

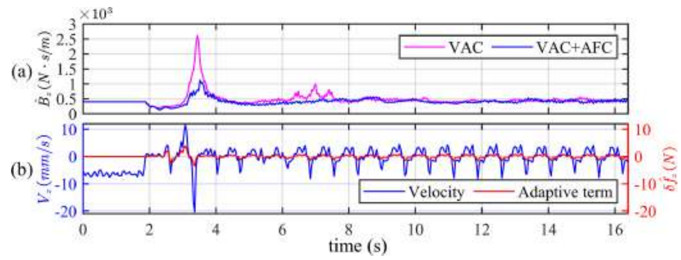

Fig. 13. Variations of the z-axis (a) damping term and (b) adaptive term during point inspection.

TABLE IV
PIPE THICKNESS INSPECTION AT 1.0 M/S WATER FLOW

	Steel pipe (mm)			PVC pipe (mm)	
Nominal thickness	1.8	2.8	3.8	6.0	9.0
Actual thickness	1.81	2.84	3.84	6.05	9.12
Min	1.79	2.81	3.82	6.03	9.08
Max	1.95	2.97	3.98	6.18	9.28
Average	1.861	2.877	3.878	6.116	9.172

Fig. 12 displays the contact forces during thickness measurement under different methods and flow velocities. In static water, the traditional admittance controller fails to regulate the contact force. The model-based variable admittance controller (MB-VAC) [26], which leverages model-based prediction optimization to adjust the stiffness term of the admittance law, successfully tracks the desired contact force. Nevertheless, the proposed variable admittance controller with adaptive feedback compensation achieves superior performance, with a lower peak force (-23.9 N), faster convergence (4.16 s), and higher accuracy (-0.63 N mean error), as summarized in Table III. The variations of the damping and adaptive terms during contact are illustrated in Fig. 13, where adaptive feedback reduces damping fluctuations and facilitates faster and smoother force convergence. At higher flow velocities, contact force regulation degrades with larger overshoot and slower convergence, but the proposed controller maintains accurate force tracking by ensuring convergence through variable damping and suppressing oscillations with adaptive compensation. Table IV presents the statistical results of the pipe thickness measurements at 1.0 m/s water flow, demonstrating that the system can measure thickness with high accuracy.

The surface cleaning trajectory in Fig. 14 is divided into three segments, showing deviations between the actual and desired paths. In the YZ plane, a 7.3° offset in the second segment indicates a rotational shift of the actual pipe relative

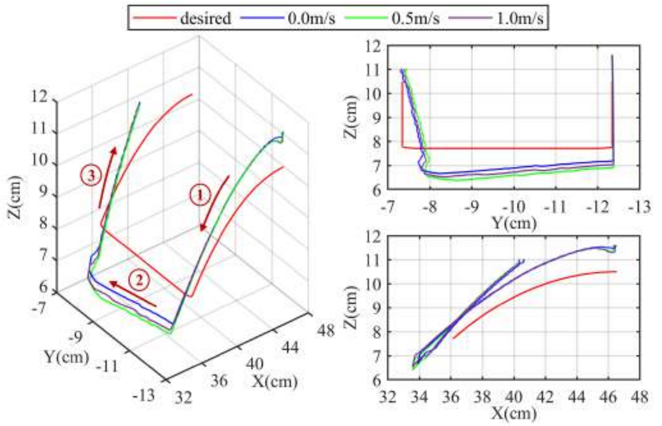

Fig. 14. Trajectories during surface cleaning using the proposed controller at different flow velocities.

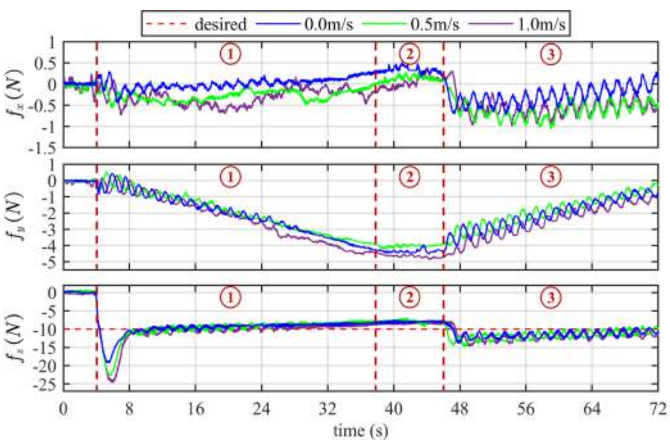

Fig. 15. Comparison of contact forces during surface cleaning using the proposed controller at different flow velocities.

to its estimation. In the XZ plane, the desired trajectory lies below the actual trajectory, with an initial gap of 1.2 cm. These deviations occur beacause the planned trajectory is based on the pipe point cloud estimation with a smaller pipe diameter (19.5 cm) and lower height than the actual pipe. Such errors highlight the necessity of compliance control to ensure proper task execution. Fig. 15 shows the contact forces under the proposed method. The z-axis overshoot is greatly reduced compared with point inspection, measuring 19.2 N at 0.0 m/s flow and 24.8 N at 1.0 m/s flow, due to the inherent compliance of the brush. Continuous brush rotation causes force fluctuations. The y-axis force varies as the brush moves along the X direction of the world frame, due to pipe estimation errors and z-axis force coupling. With point cloud guidance and compliance control, the system safely and effectively completes the curved surface cleaning task.

VI. Conclusion

This study presents an integrated underwater manipulator system for hydraulic structure inspection and maintenance, combining stereo-based perception, robust motion control, and compliant interaction. On public benchmarks, the adapted UWNet achieves 3.4% D1_all on KITTI2015 and 0.694 m RMSE on FLSea, demonstrating strong cross-domain generalization and effective domain adaptation. The robust controller, consisting of an inner-loop position controller and an outer-loop variable admittance controller, achieves a trajectory tracking RMSE of 1.44 mm at 1.0 m/s flow, corresponding to a 75.9% improvement over the PD baseline. Compared with the traditional variable admittance controller, the proposed approach effectively reduces contact impact and enhances force regulation, achieving a peak contact force of -23.9 N with a settling time of 4.16 s in still water, thereby enabling stable and safe contact interaction with minimal overshoot and oscillations and confirming its safety and reliability for underwater thickness measurement and surface cleaning tasks. Future work will investigate contact force estimation from joint torques and the integration of shared control strategies into underwater teleoperation to enhance robustness in complex and uncertain environments.

References

[1] P. Ridao, M. Carreras, D. Ribas, P. J. Sanz, and G. Oliver, “Intervention AUVs: the next challenge,” *Annu. Rev. Control*, vol. 40, pp. 227–241, 2015.

[2] A. Sans-Muntadas, M. B. Skaldebø, M. C. Nielsen, and I. Schjølberg, “Unsupervised domain transfer for task automation in unmanned underwater vehicle intervention operations,” *IEEE J. Ocean. Eng.*, vol. 47, no. 2, pp. 312–321, 2022.

[3] E. Simetti *et al.*, “Sea mining exploration with an UVMS: Experimental validation of the control and perception framework,” *IEEE/ASME Trans. Mechatronics*, vol. 26, no. 3, pp. 1635–1645, 2020.

[4] J. Mao, J. Zhou, F. Xie, Y. Liu, G. Song, and A. Song, “UTC-RS: An underwater tracked cleaning robot system for hydraulic structures,” *IEEE Robot. Automat. Lett.*, vol. 10, no. 3, pp. 2311–2318, 2025.

[5] A. Astolfi *et al.*, “Marine sediment sampling with an underwater legged robot: A user-driven sampling approach for microplastic analysis,” *IEEE Robot. Autom. Mag.*, vol. 31, no. 1, pp. 62–71, 2024.

[6] E. Simetti *et al.*, “Autonomous underwater intervention: Experimental results of the MARIS project,” *IEEE J. Ocean. Eng.*, vol. 43, no. 3, pp. 620–639, 2018.

[7] A. Birk *et al.*, “Dexterous underwater manipulation from onshore locations: Streamlining efficiencies for remotely operated underwater vehicles,” *IEEE Robot. Autom. Mag.*, vol. 25, no. 4, pp. 24–33, 2018.

[8] F. Huang, X. Yang, D. Mei, and Z. Chen, “Unified contact model and hybrid motion/force control for teleoperated manipulation in unknown environments,” *IEEE/ASME Trans. Mechatronics*, vol. 30, no. 2, pp. 921–932, 2025.

[9] J. Zhang *et al.*, “A novel handheld bronchoscope robot with human–robot shared control strategy for pulmonary examination,” *IEEE/ASME Trans. Mechatronics*, pp. 1–11, 2024.

[10] F. Huang, W. Li, Y. Wang, D. Mei, and Z. Chen, “A shared control framework with constrained optimization and guiding force feedback for teleoperation of underwater electric manipulator,” *IEEE J. Ocean. Eng.*, vol. 50, no. 2, pp. 1030–1044, 2025.

[11] F. Bonin-Font, G. Oliver, S. Wirth, M. Massot, P. Lluis Negre, and J.-P. Beltran, “Visual sensing for autonomous underwater exploration and intervention tasks,” *Ocean Eng.*, vol. 93, pp. 25–44, 2015.

[12] D. Ribas *et al.*, “I-AUV mechatronics integration for the TRIDENT FP7 project,” *IEEE/ASME Trans. Mechatronics*, vol. 20, no. 5, pp. 2583–2592, 2015.

[13] N. Palomeras *et al.*, “Toward persistent autonomous intervention in a subsea panel,” *Auton. Robots*, vol. 40, pp. 1279–1306, 2016.

[14] A. Peñalver *et al.*, “Visually-guided manipulation techniques for robotic autonomous underwater panel interventions,” *Annu. Rev. Control*, vol. 40, pp. 201–211, 2015.

[15] P. Cieślak and P. Ridao, “Adaptive admittance control in task-priority framework for contact force control in autonomous underwater floating manipulation,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, 2018, pp. 6646–6651.

[16] T. Gao, J. Yan, X. Yang, C. Chen, and X. Guan, “Integrated target recognition and grasp control for the underwater vehicle–manipulator system,” *IEEE Trans. Ind. Electron.*, vol. 71, no. 10, pp. 12 732–12 743, 2024.

[17] X. Guo, K. Yang, W. Yang, X. Wang, and H. Li, “Group-wise correlation stereo network,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 3273–3282.

[18] Z. Shen, Y. Dai, and Z. Rao, “CFNet: Cascade and fused cost volume for robust stereo matching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 13 906–13 915.

[19] Z. Shen, X. Song, Y. Dai, D. Zhou, Z. Rao, and L. Zhang, “Digging into uncertainty-based pseudo-label for robust stereo matching,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14 301–14 320, 2023.

[20] R. Gabl *et al.*, “Hydrodynamic loads on a restrained ROV under waves and current,” *Ocean Eng.*, vol. 234, p. 109279, 2021.

[21] J. Lv *et al.*, “Disturbance rejection control for underwater free-floating manipulation,” *IEEE/ASME Trans. Mechatronics*, vol. 27, no. 5, pp. 3742–3750, 2022.

[22] J. Mao, G. Song, S. Hao, M. Zhang, and A. Song, “Development of a lightweight underwater manipulator for delicate structural repair operations,” *IEEE Robot. Automat. Lett.*, vol. 8, no. 10, pp. 6563–6570, 2023.

[23] J. Na, B. Jing, Y. Huang, G. Gao, and C. Zhang, “Unknown system dynamics estimator for motion control of nonlinear robotic systems,” *IEEE Trans. Ind. Electron.*, vol. 67, no. 5, pp. 3850–3859, 2020.

[24] X. Jia, J. Yang, T. Shi, W. Wang, Y. Pan, and H. Yu, “Robust precision motion control based on enhanced unknown system dynamics estimator for high-dof robot manipulators,” *IEEE/ASME Trans. Mechatronics*, vol. 30, no. 1, pp. 357–368, 2025.

[25] S. Zhang, S. Cheng, and Z. Jin, “Variable trajectory impedance: A super-twisting sliding mode control method for mobile manipulator based on identification model,” *IEEE Trans. Ind. Electron.*, pp. 1–10, 2024.

[26] C. Seo, H. Kim, H. Jin, T. Kim, and T. Seo, “Force control of a grinding robotic manipulator with floating base via model prediction optimization control,” *IEEE/ASME Trans. Mechatronics*, vol. 28, no. 4, pp. 1911–1919, 2023.

[27] P. Di Lillo, E. Simetti, F. Wanderlingh, G. Casalino, and G. Antonelli, “Underwater intervention with remote supervision via satellite communication: Developed control architecture and experimental results within the dexrov project,” *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 1, pp. 108–123, 2020.

[28] C. Barbalata, M. W. Dunnigan, and Y. Petillot, “Position/force operational space control for underwater manipulation,” *Robot. Auto. Syst.*, vol. 100, pp. 150–159, 2018.

[29] F. Wang, H. Liu, L. Cao, F. Xie, G. Song, and A. Song, “A novel lightweight underwater manipulator based on ros2 for reliable intervention,” in *Proc. Int. Conf. Autom., Control Robot. Eng. (CACRE)*, 2024, pp. 345–349.

[30] Y. Yang and S. Soatto, “FDA: Fourier domain adaptation for semantic segmentation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 4085–4095.

[31] N. Mayer *et al.*, “A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 4040–4048.

[32] K. He, X. Zhang, S. Ren, and J. Sun, “Spatial pyramid pooling in deep convolutional networks for visual recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, 2015.

[33] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[34] Y. Randall and T. Treibitz, “FLSea: Underwater visual-inertial and stereo-vision forward-looking datasets,” *arXiv:2302.12772*, 2023.

[35] T. Guan, C. Wang, and Y.-H. Liu, “Neural markov random field for stereo matching,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2024, pp. 5459–5469.

[36] B. Wen, M. Trepte, J. Aribido, J. Kautz, O. Gallo, and S. Birchfield, “FoundationStereo: Zero-shot stereo matching,” *arXiv:2501.09898*, 2025.

Fei Wang received the M.S. degree in mechanical engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2021. He is currently pursuing the Ph.D. degree in instrument science and technology at Southeast University, Nanjing, China.

His research interests include inspection robots, underwater manipulators, compliance control, and robot learning.

HaiXin Liu received the B.S. degree in automation science from Hohai University, Nanjing, China, in 2022. She is currently pursuing the M.S. degree in instrument science and technology at Southeast University, Nanjing, China.

Her research interests include underwater manipulators and computer vision.

Cao Lin receives the B.S. degree in mechanical engineering from Wuhan University of Technology, Wuhan, China, in 2023. He is currently pursuing the M.S. degree in instrument science and technology at Southeast University, Nanjing, China.

His research interests include underwater robots and robot control.

Guangming Song (S’04-M’05-SM’12) received the Ph.D. degree in control science and engineering from the University of Science and Technology of China, Hefei, China, in 2004.

From 2004 to 2006, He was a Research Fellow with the Robotic Sensor and Control Laboratory, Southeast University, China. Since 2006, he has been with the School of Instrument Science and Engineering, Southeast University, China. He is currently a Professor with the School of Instrument Science and Engineering, Southeast University, China. His current research interests include distributed robots, aerial manipulators, and bio-inspired legged robots.

Aiguo Song (M’98-SM’11) received the B.S. degree in automatic control, the M.S. degree in measurement and control from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1990 and 1993, respectively, and the Ph.D. degree in measurement and control from Southeast University, Nanjing, China, in 1996.

He is currently a Professor and a Doctoral Supervisor with the School of Instrument Science and Engineering, Southeast University, and also the Director of the Robot Sensing and Control Technology Laboratory. His research interests include haptic display, robot tactile sensor, rehabilitation engineering, and robotic teleoperation technology. He won the National Science Fund for Distinguished Young Scholars in 2013 and the second prize of the National Technology Invention Award of China in 2017.

Roberto Meattini (Member, IEEE) received the M.Sc. degree in automation engineering and Ph.D. degree in biomedical, electrical, and systems engineering from the University of Bologna, Bologna, Italy, in 2014 and 2018, respectively. He is currently a Junior Assistant Professor with the Department of Electrical, Electronic, and Information Engineering, University of Bologna. His research interests mainly include dexterous manipulation, telemanipulation, human-in-the-loop control of robots, and human-robot interaction.

Claudio Melchiorri (Life Fellow, IEEE) received the Laurea and Ph.D. degrees in electrical engineering from the University of Bologna, Bologna, Italy, in 1985 and 1990, respectively. He is currently a Full Professor of robotics with the Department of Electrical, Electronic and Information Engineering “Guglielmo Marconi,” University of Bologna. He was with the organization of several international conferences. He has authored or coauthored about 350 scientific papers and is author of two books on trajectory planning for automatic machines and robots, and author or (co-)editor of seven books in robotics. His research interests include dexterous manipulation, haptics, telemanipulation, nonlinear control, and motion control.