Real-time Acoustic Holography with Iterative Unsupervised Learning for Acoustic Robotic Manipulation

Chengxi Zhong, Zhenhuan Sun, Teng Li, Hu Su*, and Song Liu*, *Member, IEEE*

***Abstract*— Phase-only acoustic holography is a fundamental and promising technique for contactless robotic manipulation. Through independently controlling phase-only hologram (POH) of phase array of transducers (PAT) and simultaneously driving each channel by sophisticated circuits, a certain acoustic field is dynamically generated in working medium (e.g., air, water or biological tissues) at certain moment. The phase profile of PAT is required dynamically and precisely as per arbitrary expected acoustic field for the sake of versatile and stable robotic manipulation. However, the most conventional methods rely on iterative optimization algorithms which are inevitably time-consuming and probably non-convergent, moreover hindering versatility and fidelity of acoustic robotic manipulation. To address these issues, this paper reports a real-time phase-only acoustic holography algorithm by virtue of iterative unsupervised learning. Using a physics model to construct two queues, which we refer to as experience pools, data pairs consisting of a target acoustic amplitude hologram in expected acoustic field and corresponding POH of PAT are collected on-the-fly, circumventing costly preparation of annotated dataset in advance. With iterative learning between neural network training and experience pools update, both the solution of objective inverse mapping and the adaptation for arbitrary desired acoustic field are mutually enhanced. The experiments and results validated that the proposed approach surpasses previous algorithms in terms of real time and precision.**

I. Introduction

Robotic manipulation has long been a fascinating research topic in academia. From perspective of manipulation mode and handling object's scale, it has steadily progressed from traditional contact millimetric robotic manipulation [1] to advanced contactless micro-nano robotic manipulation [2]. Every progress in robotic manipulation has further expanded the outreach of these techniques to broader related fields. For example, the optical tweezers have opened dexterous procedure of DNA measurement [3] and protein elasticity testing [4], the acoustic micromanipulation technologies have fostered noninvasive surgery in vivo [5], exogenous material delivery [6] and clinical interventions [7]. Since acoustic mechanism is imprinted with incomparably salient merits of low power consumption, favorable biocompatibility, deep penetration depth, and flexible maneuverability, acoustic contactless micro-nano robotic manipulation technique is certainly advanced and promising for various applications.

Over the past few decades, diverse methods for acoustic contactless manipulation were studied extensively in acoustic community and robotics community encompassing standing wave based acoustic levitator [8-10] and travelling wave based acoustic tweezers [11-13]. A potential acoustic field and incidental acoustic radiation force (ARF) are created due to the intricate inference of acoustic waves emitted from transducers, which permits objects trapping, levitation, locomotion, translation, and rotation, etc. Double-sided arrangement [10], which consists of an acoustic source array and a reflector or two opposite acoustic source arrays, is the most commonly used for standing wave based acoustic levitator. In such manner, nodes and antinodes generated by standing waves are employed to trap particles statically and stably, while the rotation and translation are limited. Travelling wave based holographic acoustic tweezer (HAT) [14] is an alternative with better maneuvering flexibility in which customized ARF is harvested by controlling the amplitudes and/or phase delays of a single-sided transducer array actively and individually. Recent advance [14] has confirmed that POH-based HAT allows multi-particle in-plane simultaneous manipulation and orientation, individual particles manipulation as well as 3D dynamic manipulation.

Acoustic holography [15] provides an efficient technique supporting realization of HAT for dynamic contactless micro-nano robotic manipulation in real time. Through encoding complete information including phase and amplitude of desired acoustic wavefront into a 2D hologram, the desired acoustic field is allowed to be reconstructed by motivating PAT by coherent sources referring to the 2D hologram. Thus far, both phase and amplitude modulation like GS-PAT [16], BFGS optimizer [17] and Eigensolver [18] as well as phase only modulation like LMA [19], IB [14] and Diff-PAT [20] have been broadly studied. Modulation on each element of POH by virtue of controlling delay law is easy and practical to produce steerable and focused beams. Therefore, authors focus on the study of POH-based acoustic holography. Two state-of-the-art phase-only modulation approaches are 3D printed acoustic hologram lens [21] and holographic acoustic transducer array [14]. 3D printed acoustic hologram lens records acoustic field information by pixel-wise thickness of printed phase plate. It allows high-fidelity and high-resolution acoustic wavefront reconstruction which is only limited by the 3D printed hardware theoretically. However, since it needs to be printed in advance, this technique is just suitable for static application scenarios. In contrast, holographic acoustic transducer array (HATA) is a substitute which enables dynamically objects manipulation through individually

*This paper was sponsored by Shanghai Pujiang Program under Grant 21PJ1410500, and in part by the National Natural Science Foundation of China under Grant 61906191. (*Corresponding Author: Hu Su and Song Liu*)

C. Zhong, Z. Sun, and T. Li are with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China (e-mail: zhongchx, sunzhh, liteng1@shanghaitech.edu.cn).

H. Su is with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: hu.su@ia.ac.cn).

S. Liu is with the School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China, and with Shanghai Engineering Research Center of Intelligent Vision and Imaging, Shanghai, China (e-mail: liusong@shanghaitech.edu.cn).

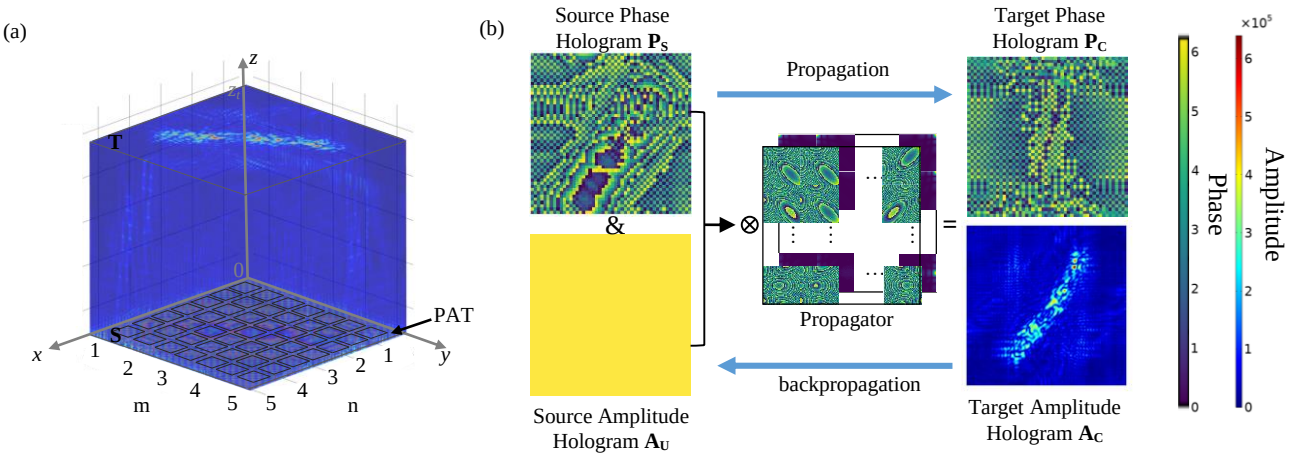

Fig. 1. Schematic of inverse kinematics problem. (a) shows the system model equipping by PAT. (b) shows the propagation and backpropagation between source holograms and target holograms.

modulating each channel of PAT based on computed phase profile as per acoustic transducer array configuration and desired acoustic field distribution. Therefore, HATA provides an effective technique to realize dynamic and precise non-contact micro-nano robotic manipulation in real time.

Intuitively, precise and real-time computation of phase holograms are necessary, which directly determines the reconstruction quality of acoustic field and highly influences the stability and versatility of holographic acoustic manipulation. Conventional methods for POH calculation depend on iterative optimization, such as, Gerchberg Saxton (GS) algorithm [22], Iterative Angular Spectrum Approach (IASA) [21] and Iterative Backpropagation (IB) algorithm [14]. However, these methods are time-consuming and less accurate, especially for acoustic field which are characterized by large number of focal points and complicated holographic modes. Recently, deep learning offers a prospect of calculating POHs as per expected acoustic field rapidly and precisely [23-27]. Regardless of neural network (NN) training process, POH for PAT are allowed to be inferred by means of just once forward calculation within a few milliseconds which merely depends on the complexity of network architecture.

In this paper, we propose an iterative unsupervised learning algorithm with a physics model named angular spectrum method (ASM) [21], using U-Net [28] architecture. The proposed algorithm bypasses costly annotated dataset acquisition in advance. Specifically, inspired by reinforcement learning [29], we employ two queues, which is referred to as experience pools, to collect and update data pairs during training process to regress mapping and alleviate domain bias. As the network steadily reaches the optimized solution during iterative training process, the periodically collected experience pools simultaneously gets closer to the desired acoustic field distribution, which further refines the network's understanding of the objective inverse kinematics problem. Additionally, self-supervised learning (SSL) with help of physical model ASM, is employed in order to accelerate convergence of NN. During the iterative network learning and periodic experience pools update, they are mutually enhanced so as to optimize the objective inverse mapping for arbitrary desired acoustic field reconstruction task. The experiments and results evaluate that employed experiment pools effectively leverage network convergence and domain adaptation. Moreover, it is evidently confirmed that our proposed acoustic holography approach surpasses the existing iterative optimization methods with time consumption of 131 milliseconds implemented on GPU, average peak-signal-noise-ratio (PSNR) of 17.69 and average mean square error (MSE) of 0.017. Furthermore, it isn't limited by customized arrangement of phase-only transducer array and tailored distribution of holographic acoustic field theoretically.

II. Inverse Kinematics Problem

By referring our previous work [30], the studied inverse kinematics problem is succinctly discussed here. As illustrated in Fig. 1 (a), the Cartesian coordinates (x, y, z) with origin sitting on the left corner of PAT are established. The PAT consists of $m \times n$ independent transducers laying on $z = z_0 = 0$ plane, which acts as acoustic source denoted as $\mathbf{S} \in \mathbb{C}^{m \times n}$. The target wavefront of interest is interpreted as $m \times n$ acoustic hologram, laying on the $z = z_t$ plane denoted as $\mathbf{T} \in \mathbb{C}^{m \times n}$, which parallels with acoustic source plane. $\mathbb{C}$ refers to the complex field including two physical degrees of freedom, i.e., amplitude and phase information of acoustic waves. The identical resolution between source plane and target hologram is due to the limitation of spatial bandwidth product. The complex acoustic pressures p of $\mathbf{S}$ and $\mathbf{T}$ can be expressed as

$$p(x, y, z) = A(x, y, z) e^{j\varphi(x, y, z)} \tag{1}$$

where $p(x, y, z)$ is acoustic pressure, $A(x, y, z)$ is amplitude, and $\varphi(x, y, z)$ is phase at position (x, y, z) on $\mathbf{S}$ or $\mathbf{T}$ plane. ASM is a physics model formulating the acoustic field propagation, which is expressed as

$$\begin{cases} p(x, y, z_t) = \mathrm{IFFT}\left\{\mathrm{FFT}\left\{p(x, y, z_0)\right\} * H\left(k_x, k_y, z\right)\right\} \\ H\left(k_x, k_y, d\right) = exp\left[-jd\sqrt{\left(k^2 - k_x^{\,2} - k_y^{\,2}\right)}\right] \end{cases} \tag{2}$$

where FFT and IFFT are Fast Fourier Transform and Inverse Fast Fourier Transform, respectively, d refers to propagation distance, $k = (k_x, k_y, k_z)$ is the wave vector, k_x, k_y are its x and y components, and $H \in \mathbb{C}^{m \times n \times m \times n}$ is the propagator which theoretically expresses the amplitude and phase changes of acoustic waves in frequency domain during the propagation.

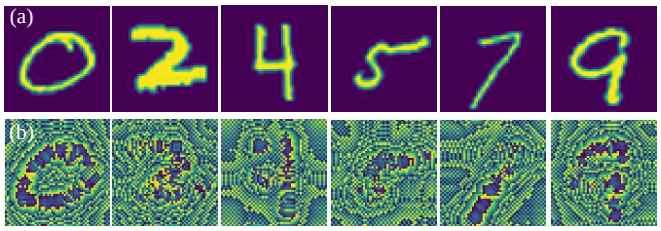

Fig. 2. (a) shows the examples of expected target amplitude hologram $\mathbf{A_E}$ and (b) shows corresponding source phase hologram $\mathbf{P_S}$.

The $*$ is the convolution symbol. It is shown that the acoustic field propagation modelled by ASM can be interpreted as a differentiable convolution operation but in the complex field. The acoustic field backpropagation given as Eq. (3) is realized by substituting propagator H in Eq. (2) with its conjugate value. The convolution operation with propagator or its conjugate value is discriminated by setting a positive or negative propagation distance parameter.

$$\begin{cases} p(x,y,z_0) = \mathrm{IFFT}\left\{\mathrm{FFT}\left\{p(x,y,z_t)\right\} * \overline{H}\left(k_x,k_y,-z\right)\right\} \\ \overline{H}\left(k_x,k_y,-z\right) = exp\left[\, jz\sqrt{\left(k^2 - {k_x}^2 - {k_y}^2\right)}\,\right] \end{cases} \tag{3}$$

The acoustic field forward propagation and backpropagation between target hologram $\mathbf{T}$ and source hologram $\mathbf{S}$ is illustrated by Fig. 1 (b). From perspective of robotics, the acoustic field forward propagation can be regarded as forward kinematics denoted as $\boldsymbol{f}$, while the acoustic field backpropagation can be viewed as inverse kinematics denoted as $\boldsymbol{f}^{\dagger}$. Intuitively, instead of solving simple and linear forward kinematics represented as Eq. (4), the inverse kinematics problem is mathematically non-linear expressed in Eq. (5).

$$\boldsymbol{f} : \mathbf{S}(x,y,z_0) \mapsto \mathbf{T}(x,y,z_t),\ where\ x \le m, y \le n \tag{4}$$

$$\boldsymbol{f}^{\dagger} : \mathbf{T}(x,y,z_t) \mapsto \mathbf{S}(x,y,z_0),\ where\ x \le m, y \le n \tag{5}$$

In this paper, the investigated phase-only holography problem is modelled by constraining amplitude of source hologram $\mathbf{S}$ to be uniform distributed denoted as $\mathbf{A_u} \in \mathbb{R}^{m \times n}$. Meanwhile, in order to pursue certain robotic manipulation, the amplitude of target hologram $\mathbf{T}$ is expected to extremely approximate to the tailored modality with an acceptable error, which is denoted as $\mathbf{A_E} \in \mathbb{R}^{m \times n}$. Therefore, considering the nonlinearity and mathematical unsolvability of the inverse kinematics problem, this work proposes an iterative unsupervised learning algorithm in conjunction with a physics model ASM to regress the inverse mapping from the target amplitude hologram $\mathbf{A_E}$ to the source phase hologram $\mathbf{P_S}$. Then, the tailored target amplitude hologram $\mathbf{A_E}$ is allowed to be reconstructed precisely from the computed source phase hologram $\mathbf{P_S}$ with the uniform distributed source amplitude hologram $\mathbf{A_u}$. The source phase hologram $\mathbf{P_S}$ sitting in the range of $[0, 2\pi)$ and the normalized target amplitude hologram $\mathbf{A_E}$ sitting in the range of $[0, 1]$ can be vividly viewed as grayscale images in which each pixel value represents phase $\varphi(x, y, z)$ or amplitude $A(x, y, z)$, as shown by Fig. 2.

III. Iterative Unsupervised Learning Algorithm

The data acquisition, a traditional preparation in well-known supervised learning, is laborious, time-consuming and sometimes impossible. Especially for the inverse kinematics problem investigated in this paper, neither simulation approach nor physical measurement is adequately suitable to synthesize data pairs which precisely corresponds to arbitrary expected acoustic field morphologies of interest for dexterous and complicated robotic manipulation [25]. Although the sufficiently approximated data pairs can be utilized, the challenge of domain bias would be incurred. Therefore, this work proposes an iterative unsupervised learning algorithm with a physics model, which circumvents data acquisition in advance, meanwhile, by constructing experience pools to eliminate domain bias and improve the reconstruction quality. In our proposed algorithm, four stages, including on-the-fly experiences collection, SSL-based warmup for convergence acceleration, and inverse kinematics mapping regression in two distinct experience pools, are iteratively performed. Details of the algorithm and employed techniques are articulated in this section.

A. On-the-fly Experiences Collection

Escaping from collection data pairs in advance, this work dynamically and asynchronously constructs and updates two experience pools during training process. As the module shown in top left of Fig. 3, the desired amplitude holograms $\mathbf{A_E}$ sampled from original dataset are fed into NN, outputting their predicted source POH. The predicted POH $\mathbf{P_S}$ and corresponding reconstructed amplitude holograms $\mathbf{A_C}$ are packed as a data pair and stacked into the first pool $\mathbf{Q_1}$. Preserving the reconstructed phase holograms $\mathbf{P_C}$ and utilizing $\mathbf{A_E}$, the acoustic field back propagation is conducted, whereby $\mathbf{A_E}$ and retrieved source phase hologram $\mathbf{P^S}$ are packed as a data pair and stacked into the second pool $\mathbf{Q_2}$. During the iterative training, the maintained $\mathbf{Q_1}$ and $\mathbf{Q_2}$ are enhanced, providing powerful references for inverse kinematics mapping.

The data pairs in $\mathbf{Q_1}$ satisfy phase-only modulation of PAT in our system, while the data pairs in $\mathbf{Q_2}$ is closely approximate to target domain. Due to the data pairs in experience pools might come from different iterations, the consistency of each pool and the balance between two pools are highly significant. Therefore, the experience pools are respectively equipped with pre-optimized capacities of N_1, N_2 and update frequencies of C_1, C_2. In order to further mitigate the influence caused by inconsistency, the data pairs are randomly sampled from $\mathbf{Q_1}$ and $\mathbf{Q_2}$. The applied techniques of experience pools leverage simultaneously optimizing NN and decreasing domain bias, meanwhile, preventing network from oscillation.

B. SSL-based Warmup

To provide a better initialization of network and accelerate convergence, SSL-based warmup is conducted, as the module shown in top right of Fig.3. In this stage, the reconstruction accuracy is directly penalized instead of supervising network output. In detail, the phase holograms of PAT are regressed by feeding $\mathbf{A_E}$ into network. Afterward, the acoustic field propagation is conducted using ASM on network output and $\mathbf{A_u}$. The difference between $\mathbf{A_C}$ and $\mathbf{A_E}$ is calculated by two loss items, including MSE, denoted as l_{mse} and energy loss, denoted as l_{energy}. MSE is used to penalize from pixel-wise perspective, while the energy loss is to penalize with respect to energy proportion of foreground and background in $\mathbf{A_C}$. The used loss items are expressed as

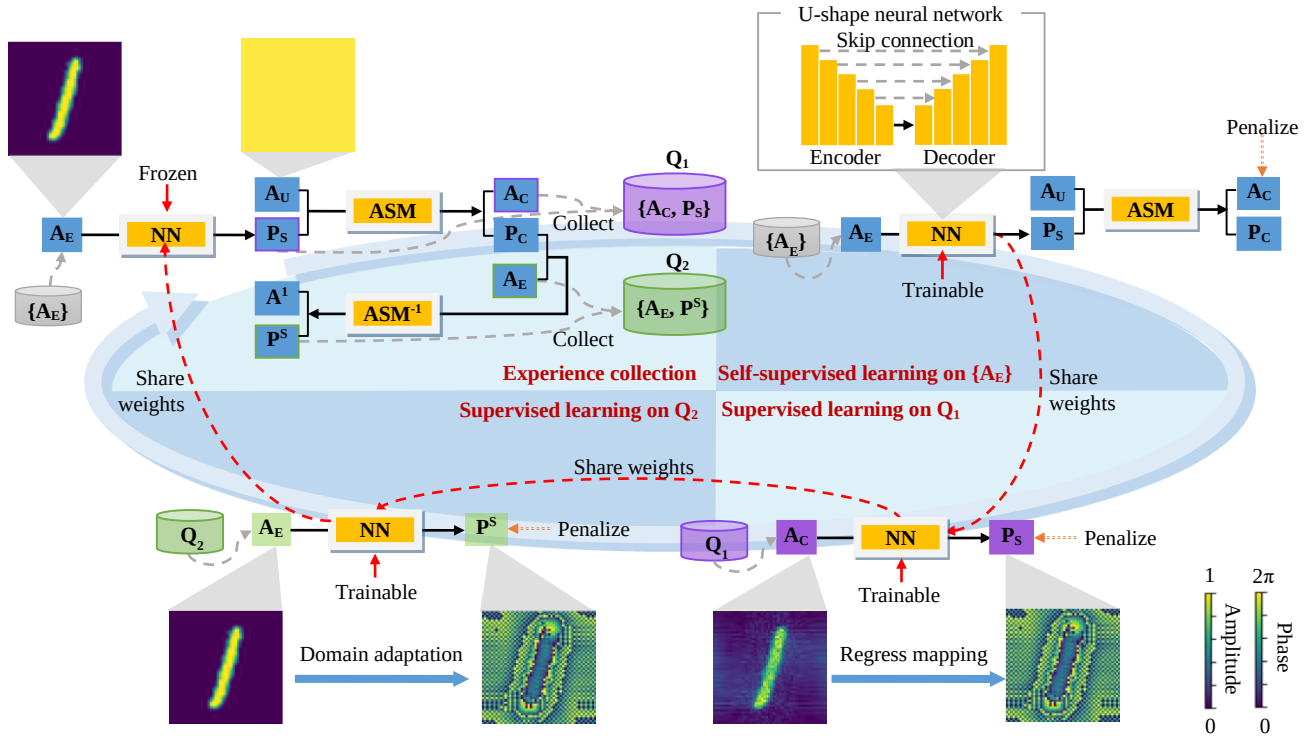

Fig. 3. Illustration of iterative unsupervised learning algorithm. It consists of four stages, i.e., experience collection, self-supervised learning on $\{\mathbf{A}_E\}$, supervised learning on $\mathbf{Q}_1$ and $\mathbf{Q}_2$. These four stages are iteratively conducted to optimize network and eliminate domain bias step by step until the acceptable reconstruction performance of expected target amplitude hologram is yielded.

$$\begin{cases} l_{mse} = \frac{1}{N}\sum\left(\mathbf{A}_C' - \mathbf{A}_E\right) \\ l_{energy} = P_{eng}^{\ B}\left(\mathbf{A}_C'\right) + \left[1 - P_{eng}^{\ F}\left(\mathbf{A}_C'\right)\right] \end{cases} \tag{6}$$

where $\mathrm{A_C}'$ is reconstructed by network output, $P_{eng}^{B}(\cdot)$ and $P_{eng}^{F}(\cdot)$ are foreground and background proportions of $\cdot$, respectively. The SSL stage plays a crucial role to warm up network for comparably better fitting the inverse kinematics mapping. After SSL-based warmup, the network weights are shared with afterward learning conducted on experience pools.

C. Inverse Kinematics Mapping Regression

The Experience pool $\mathbf{Q}_1$ composing of $\{\mathbf{A}_C, \mathbf{P}_S\}$ is used for inverse kinematics mapping regression. Since data pairs in this experience pool are in accordance with POH configuration, they are used to learning precise inverse mapping theoretically. In this stage, the previous collected splotchy target amplitude holograms $\mathbf{A}_C$ are inputted to the network and corresponding source phase holograms $\mathbf{P}_S'$ are predicted. Then, ASM is used to reconstruct target amplitude holograms. This stage is shown in bottom right of Fig. 3. Taking the periodic nature of phase into consideration [30], the penalization denoted as l_{cos} applied to update parameters of network in this stage is given as

$$l_{cos} = \frac{1}{N}\sum\left(1 - \cos\left(\mathbf{P}_S - \mathbf{P}_S'\right)\right) \tag{7}$$

D. Target Domain Adaptation

Afterward, the experience pool $\mathbf{Q}_2$ is used for domain adaptation shown as bottom left of Fig. 3. The initial network parameters of domain adaptation stage are shared from that of well-trained network in inverse kinematics regression stage. The random sampled expected target holograms $\mathbf{A}_E$ from $\mathbf{Q}_2$ are fed into network, then outputting predicted source phase holograms whose ground truth are $\mathbf{P}^S$. The used loss function is same as that given by Eq. (7). Thanks to the shared weights from previous regression stage, the network just requires lightly fine-tuning to adapt target domain in this stage. Finally, to optimize network parameters and eliminate domain bias step by step, the aforementioned four stages are iteratively conducted. The iterations will terminate until the acceptable reconstruction performance is yielded.

IV. Experiments and Results

The experimental results in this section confirms the feasibility, versatility and efficiency of the proposed iterative unsupervised learning algorithm. Firstly, the NN architecture and implementation detail are expressed. Afterward, the necessity of experience pools and the effectiveness of iterative training are confirmed. Consequently, the accuracy and real-time performance are manifested by comparing with state-of-the-art (SOTA) algorithms, including IASA, IB and GS-PAT.

A. Network Architecture and Implementation Detail

This paper proposed an iterative unsupervised learning algorithm which is in conjunction with a physics model and synthesized experience pools. The network takes the charge of solving the inverse mapping from target amplitude holograms to source POHs by the original dataset $\{\mathbf{A}_E\}$, maintained experience pools $\mathbf{Q}_1$ and $\mathbf{Q}_2$. In detail, the desired target amplitude holograms $\mathbf{A}_E$ or the reconstructed splotchy target amplitude holograms $\mathbf{A}_C$ with size of 50×50 pixels and with a reconstruction axial distance of 30 millimeters are inputted into U-Net and corresponding POHs with same size are

predicted. The U-Net [28], which consists of a contracting path and an expansive path, was applied with customized dimensions. Each path involves 4 blocks. Each block in contracting path consists of two 3×3 Convolutional layers and a 2×2 max pooling with stride of 2 for downsampling to encode features from input holograms, while upsampling layers and de-convolutional layers were repetitively employed to recover the extracted features into phase space in expansive path. Each convolutional layer or de-convolutional layer was followed by ReLU activity function and Batch Normalization. Moreover, the skip connections were used between the tailored contracting and expansive paths. Based on the output, ASM was optionally conducted for acoustic field reconstruction.

The algorithm was implemented by PyTorch on a sever with 4 Nvidia Tesla M40 GPUs, running on Ubuntu 18.04 operating system. The used MNIST dataset, in which the foreground pixels of digital number image act as focal points, support robotic manipulation. The sever was equipped with Intel Core i7-10700 CPU having the frequency of 2.90 GHz. SSL-based warmup was performed for 50 epochs. Distinct stages were iteratively performed for 20 times with shared parameters from the previous stage. Training epochs on $\mathbf{Q_1}$ and $\mathbf{Q_2}$ were 1 and 5. The capacities N_1 and N_2 of $\mathbf{Q_1}$ and $\mathbf{Q_2}$ were set as 32 and 160. The update frequencies $C_1 = C_2 = 20$. When an update is required, the latest network was used for renewing $\mathbf{Q_1}$ and $\mathbf{Q_2}$. The Adam optimizer is employed with learning rates of 0.001, 0.0001 and 0.0001 for SSL, and training on $\mathbf{Q_1}$ and $\mathbf{Q_2}$.

B. Ablation Study of Techniques and Loss Items

The constructed experience pools $\mathbf{Q_1}$ and $\mathbf{Q_2}$ make full use of data pairs generated during training process instead of laborious and costly dataset synthetization in advance. To confirm the effective enhancement of the experience pool technique, a randomly sampled example in experience pool $\mathbf{Q_1}$ at three successive iterations are illustrated in Fig.4. The steadily increased PSNRs which are shown at the right corner of Fig. 4 (a), as well as the decreased amplitude and phase differences which are highlighted by red rectangular in histogram of Fig. 4 visually verified the effective enhancement in terms of $\mathbf{Q_1}$, which intuitively leads to the improved retrieved phases in $\mathbf{Q_2}$. Apart from the qualitative evaluation, another ablation study was performed quantified by PSNR and MSE to verify the effectiveness of techniques and also loss items, as shown in Table I. With comparison among ablation studies indexed by 2, 3 and 5, the case indexed by 5 has the highest PSNR and lowest MSE, demonstrating the necessities of $\mathbf{Q_1}$ and $\mathbf{Q_2}$. From the comparison depicted among ablation studies indexed by 1, 4 and 5, the functionality and necessity of employed SSL-based warmup was testified. Moreover, we conducted ablation study to prove the necessity of each loss item in distinct training stages. L_{mse} and l_{cos} are basic loss items used in SSL and experience pools training stages, respectively. The improvement due to additional l_{energy} loss item used in SSL-based warmup was proved by the comparison between cases indexed by 4 and 5. To the end, the results shown in Table I manifested that our proposed algorithm yielded the highest PSNR of 17.69 and the lowest MSE of 0.017.

C. Acoustic Field Reconstruction Accuracy

Apart from the evaluation in term of functionality and necessity of distinct techniques and loss items used in the proposed algorithm, the reconstruction accuracy of arbitrary

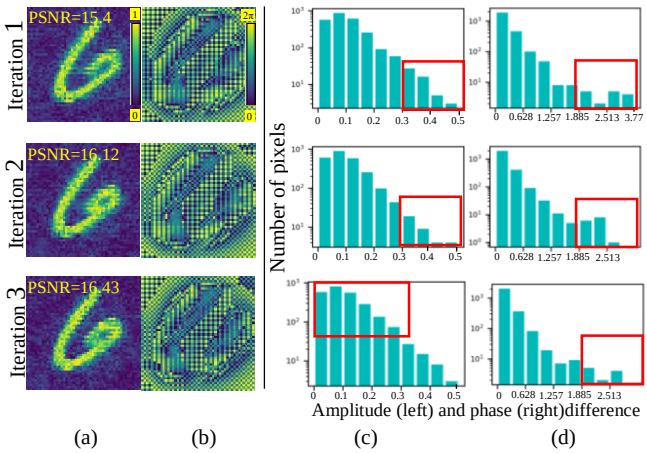

Fig. 4. Iterative enhancement examples of experience pool $\mathbf{Q_1}$. (a) is **Ac** (b) is $\mathbf{P_S}$. (c) shows corresponding histogram analysis on pixel-wise distance of $\mathbf{A_C}$ from expected target hologram $\mathbf{A_E}$ in which red rectangular is used to highlight the enhancement of pool $\mathbf{Q_1}$. (d) is the phase discrepency between experience pool $\mathbf{Q_1}$ and $\mathbf{Q_2}$.

TABLE I. ABLATION STUDY ON TECHNIQUES AND LOSS ITEMS

Index	Techniques			Loss items			PSNR ↑	MSE ↓
	SSL	$\mathbf{Q_1}$	$\mathbf{Q_2}$	l_{mse}	l_{energy}	l_{cos}		
1		√	√			√	16.03	0.025
2	√		√	√	√	√	14.89	0.033
3	√	√		√	√	√	13.95	0.040
4	√	√	√	√		√	16.42	0.022
5	√	√	√	√	√	√	**17.69**	**0.017**

desired $\mathbf{A_E}$ was evaluated qualitatively and quantitatively. For qualitative evaluation, the reconstructed results obtained by the well-trained network were depicted in Fig. 5. Fig. 5 (a) shows the $\mathbf{A_E}$ sampled from MNIST dataset, Fig. 5 (b) shows the reconstructed results from predicted POH by the network shown in Fig. 5 (d), and Fig. 5 (c) shows the difference between Fig. 5 (a) and (b). For quantitative evaluation, five perceptual or statistical metrics, including PSNR, structure similarity (SSIM), mean absolute error (MAE), accuracy and efficacy [31] were studied by comparing the optimization methods with IASA, IB and Diff-PAT. Fig.6 proves that the proposed algorithm surpassed the SOTA algorithms with improved or comparable accuracy and stability performance.

D. Comparison on Real-time Performance

Another superiority of our proposed iterative unsupervised learning algorithm is its excellent real-time performance, which was compared with IASA, IB and Diff-PAT as articulated in Table II. Thanks to that the network can just take a few of milliseconds to infer by forward inferring once it was well-trained, the average speed for predicting POH of the proposed method on both GPU and CPU is faster than that of SOTA algorithms. More significantly, the computational time is not influenced by complexity of acoustic modalities, while SOTA iterative optimization methods might not converge with the increased complexity of acoustic modalities. The network can be tailored by changing its dimensions and depth to fit different size of PAT with comparable computing time. On all account, the precise and real-time performance indicated the significance of the proposed algorithm for dynamical and dexterous non-contact micro-nano robotic manipulation.

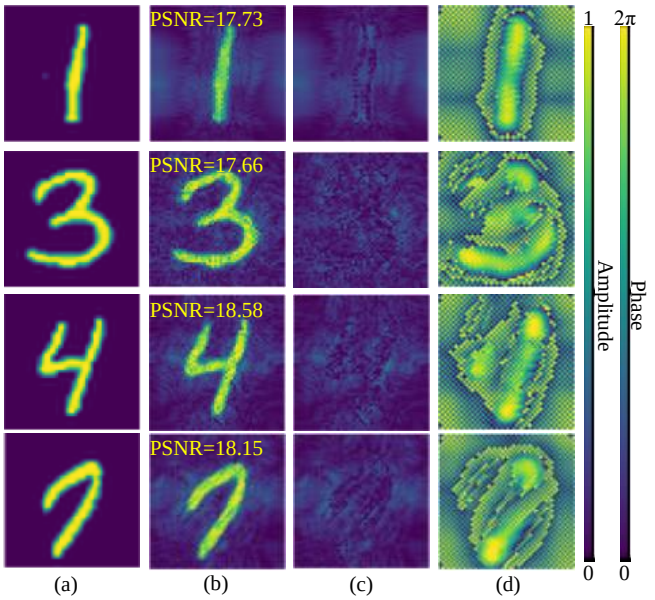

Fig. 5. Perceptual results of the proposed method. (a) shows expected target amplitude holograms $\mathbf{A}_E$, (b) is reconstructed splotchy target amplitude holograms $\mathbf{A}_C$ whose PSNR are labelled, (c) is the difference between (a) and (b), and (d) is the predicted source phase holograms outputed from network.

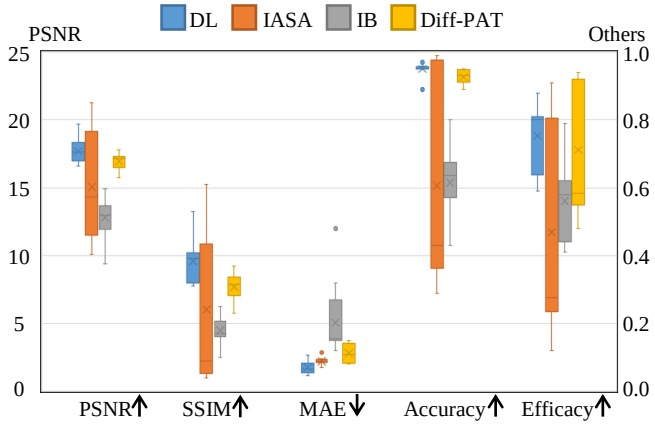

Fig. 6. Comparison in terms of distinct evaluation metrics between our proposed method and SOTA algorithms including IASA, IB and Diff-PAT.

TABLE II. COMPARISON WITH SOTA IN TERMS OF REAL TIME PERFORMANCE

Method	Proposed Method	IASA [21]	IB [14]	Diff-PAT [20]
CPU	**149 ms**	5031 ms	~30 min	11.2 s
GPU	**131 ms**	3210 ms	582 s	7.5 s

V. CONCLUSION

This paper demonstrates an iterative unsupervised learning algorithm for real time acoustic holography. By a physics model named ASM, experience pools were collected and updated during training process bypassing costly annotated dataset acquisition. With iteratively mutual enhancement between NN and experience pools, the inverse kinematics mapping from desired target amplitude holograms to the source phase holograms were figured out with a superior precise and real time performance. Therefore, the calculated source POHs support accurate acoustic field reconstruction in real time for contactless micro-nano robotic manipulation. Moreover, the used of ASM blurs the boundary between deep NN and physics community. In future research, we plan to design adequate physical experiments to incorporate the theoretical method into a real acoustic manipulation system in practice. Afterword, we will extend the applicability of the proposed method from 2D to versatile 3D acoustic reconstruction for widespread application.

REFERENCES

[1] Y. Maeda, H. Kijimoto, Y. Aiyama, and T. Arai, "Planning of Graspless Manipulation by Multiple Robot Fingers," *in Proc. IEEE Int. Conf. Robot. Autom.* (*ICRA*), 2001, pp. 2474-2479.

[2] S. Liu, Y. F. Li, and X. W. Wang, "A Novel Dual-Probe based Micro-grasping System Allowing Dexterous 3D Orientation Adjustment," *IEEE Trans. Autom. Sci. and Eng.*, vol. 17, no. 4, pp. 2048-2062, May 2020.

[3] C.J. Bustamante, Y.R. Chemla, S. Liu, and M.D. Wang, "Optical tweezers in single-molecule biophysics," *Nature Reviews Methods Primers*, vol. 1, no. 1, pp. 1-29, Mar 2021.

[4] N. Rezaei, B.P. Downing, A. Wieczorek, C.K. Chan, R.L. Welch, and N.R. Forde, "Using optical tweezers to study mechanical properties of collagen," In *Photonics North*, vol. 8007, pp. 146-155, Aug 2011.

[5] M. A. Ghanem, A. D. Maxwell, Y. N. Wang, B. W. Cunitz, V. A. Khokhlova, O. A. Sapozhnikov and M.R. Bailey, "Noninvasive acoustic manipulation of objects in a living body," *Proceedings of the National Academy of Sciences* (*PNAS*), vol. 117, no. 29, pp, 16848-16855, Jul 2020.

[6] S. Yoon, M.G. Kim, C.T. Chiu, J.Y. Hwang, H.H. Kim, Y. Wang, and K.K. Shung, "Direct and sustained intracellular delivery of exogenous molecules using acoustic-transfection with high frequency ultrasound," *Scientific reports*, vol. 6, no. 1, pp. 1-11, Feb 2016.

[7] M. Wiklund, R. Green, and M. Ohlin, "Acoustofluidics 14: Applications of acoustic streaming in microfluidic devices," *Lab on a Chip*, vol. 12, no. 14, pp. 2438-2451, 2012.

[8] P. Glynne-Jones, C.E. Demore, C. Ye, Y. Qiu, S. Cochran, and M. Hill, "Array-controlled ultrasonic manipulation of particles in planar acoustic resonator," *IEEE Trans. ultrasonics, ferroelectrics, and frequency control*, vol. 59, no. 6, pp. 1258-1266, Jun 2012.

[9] J. Lee, S.Y. Teh, A. Lee, H.H. Kim, C. Lee, and K.K. Shung, "Single beam acoustic trapping," *Applied physics letters*, vol. 95, no. 7, pp. 073701, Aug 2009.

[10] A. Marzo, S. A. Seah, B. W. Drinkwater, D. R. Sahoo, B. Long, and S. Subramanian, "Holographic acoustic elements for manipulation of levitated objects," *Nature communications*, vol. 6, no. 1, pp. 1-7, Oct. 2015.

[11] Y. Ochiai, T. Hoshi, and J.R. ekimoto, 2014." Pixie dust: graphics generated by levitated and animated objects in computational acoustic-potential field," *ACM Trans. Graphics* (*TOG*), vol. 33, no. 4, pp. 1-13, Jul 2014.

[12] L. Meng, F. Cai, F. Li, W. Zhou, L. Niu and H. Zheng, "Acoustic Tweezers," *Journal of Physics D: Applied Physics*, vol. 52, no. 27, pp. 273001, May 2019.

[13] J. Li, A. Crivoi, X. Peng, L. Shen, Y. Pu, Z. Fan, and S.A. Cummer, "Three dimensional acoustic tweezers with vortex streaming," *Communications Physics*, vol. 4, no. 1, pp. 1-8, Jun 2021.

[14] A. Marzo, and B. W. Drinkwater, "Holographic Acoustic Tweezers," *Proceedings of the National Academy of Sciences* (*PNAS*), vol. 116, no. 1, pp. 84-89, Jan. 2019.

[15] J.D. Maynard, E.G., Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," *The Journal of the Acoustical Society of America*, vol. 78, no. 4, pp. 1395-1413, Oct 1985.

[16] D.M. Plasencia, R. Hirayama, R. Montano-Murillo, and S. Subramanian, "GS-PAT: High-speed Multi-point Sound-fields for Phased Arrays of Transducers," *ACM Trans. Graphics*, vol. 39, no. 4, pp. 138-1, Jul 2020.

[17] R. Hirayama, G. Christopoulos, D. Martinez Plasencia, and S. Subramanian, "High-speed acoustic holography with arbitrary

scattering objects," *Science advances*, vol. 8, no. 24, pp .eabn7614, Jun 2022.

[18] B. Long, S.A. Seah, T. Carter, and S. Subramanian, "Rendering volumetric haptic shapes in mid-air using ultrasound," *ACM Trans. Graphics*, vol. 33, no. 6, pp.1-10, Nov 2014.

[19] E. Sakiyama, A. Matsubayashi, D. Matsumoto, M. Fujiwara, Y. Makino, and H. Shinoda, "Midair tactile reproduction of real objects,". In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*, pp. 425-433, Sep 2020.

[20] T. Fushimi, K. Yamamoto and Y. Ochiai, "Acoustic hologram optimisation using automatic differentiation," Scientific reports, vol. 11, no. 1, pp. 1-10, Jun. 2021.

[21] K. Melde, A. G. Mark, T. Qiu, and P. Fischer, "Holograms for acoustics," *Nature*, vol. 537, no. 7621, pp. 518-522, Sep. 2016.

[22] R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, no. 2, pp. 237-246, 1972.

[23] Q. Lin, J. Wang, F. Cai, R. Zhang, D. Zhao, X. Xia, J. Wang, and H. Zheng, "A deep learning approach for the fast generation of acoustic holograms", *The Journal of the Acoustical Society of America*, vol. 149, no. 4, pp. 2312-2322, Apr. 2021.

[24] S. C. Liu, and D. Chu, "Deep learning for hologram generation," *Optics Express*, vol. 29, no. 17, pp. 27373-27395, Aug. 2021.

[25] C. Zhong, Y. Jia, D. C. Jeong, Y. Guo, and S. Liu, "AcousNet: A deep learning based approach to dynamic 3D holographic acoustic field generation from phased transducer array," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 666-673, Nov. 2021.

[26] H. Goi, K. Komuro, and T. Nomura, "Deep-learning-based binary hologram," *Applied Optics*, vol. 59, no. 23, pp. 7103-7108, Aug 2020.

[27] L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks," *Nature*, vol. 591, no. 7849, pp. 234-239, Mar. 2021.

[28] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241, Springer, Oct 2015.

[29] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, and S. Petersen, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529-533, Feb 2015.

[30] C. Zhong, Z. Sun, K. Lv, Y. Guo, and S. Liu, "Real-time Acoustic Holography with Physics-based Deep Learning for Acoustic Robot Manipulation," *IEEE International Conference on Intelligent Robots and Systems* (*IROS 2022*), Kyoto, Japan, Oct. 23-27, 2022.

[31] M.H. Eybposh, N.W. Caira, P. Chakravarthula, M. Atisa, and N.C. Pégard, "High-speed computer-generated holography using convolutional neural networks," in *Optics and the Brain*, pp. BTu2C-Optica Publishing Group, April 22020.