

# Digital holographic microscopy applied to 3D computer micro-vision by using deep neural networks

Stéphane Cuenat<sup>a</sup> , Jesús E. Brito Carcaño<sup>a</sup>, Belal Ahmad, Patrick Sandoz , Raphaël Couturier ,  
Guillaume J. Laurent, and Maxime Jacquot\* 

Université de Franche-Comté, SUPMICROTECH-ENSMM, CNRS, Institut FEMTO-ST, 25000 Besançon, France

Received 31 January 2024 / Accepted 13 June 2024

**Abstract.** Deep neural networks (DNNs) are increasingly employed across diverse fields of applied science, particularly in areas like computer vision and image processing, where they enhance the performance of instruments. Various advanced coherent imaging techniques, including digital holography, leverage different deep architectures like convolutional neural networks (CNN) or Vision Transformers (ViT). These architectures enable the extraction of diverse metrics such as autofocusing reconstruction distance or 3D position determination, facilitating applications in automated microscopy and phase image restitution. In this work, we propose a hybrid approach utilizing an adapted version of the GedankenNet model, coupled with a UNet-like model, for the purpose of accessing micro-objects 3D pose measurements. These networks are trained on simulated holographic datasets. Our approach achieves an accuracy of 98% in inferring the 3D poses. We show that a GedankenNet can be used as a regression tool and is faster than a Tiny-ViT (TViT) model. Overall, integrating deep neural networks into digital holographic microscopy and 3D computer micro-vision holds the promise of significantly enhancing the robustness and processing speed of holograms for precise 3D position inference and control, particularly in micro-robotics applications.

**Keywords:** Digital holography, Microscopy, Computer micro-vision, Deep neural networks.

## 1 Introduction

In computer vision and robotics, accurate 3D positioning and trajectory determination are crucial for a variety of applications, including industrial and clinical [1]. Neural networks, including convolutional neural networks (CNNs) or Vision Transformers (ViT) play a significant role in visual data processing [2]. Digital holography (DH) in microscopy enhances the analysis of object amplitude and phase in a single image with off-axis configuration, improving the accuracy of in-focus position detection without mechanical adjustments. Combining Deep Neural Networks (DNN), mixing version of the GedankenNet model [3] and a UNet-like model [4] with DH provides a promising solution for accurately controlling complex trajectories of micro-objects in automated microscopy in real-time constrains [5].

<sup>a</sup>Jesús E. Brito Carcaño and Stéphane Cuenat contributed equally to this work as first authors.

\* Corresponding author: [maxime.jacquot@univ-fcomte.fr](mailto:maxime.jacquot@univ-fcomte.fr)

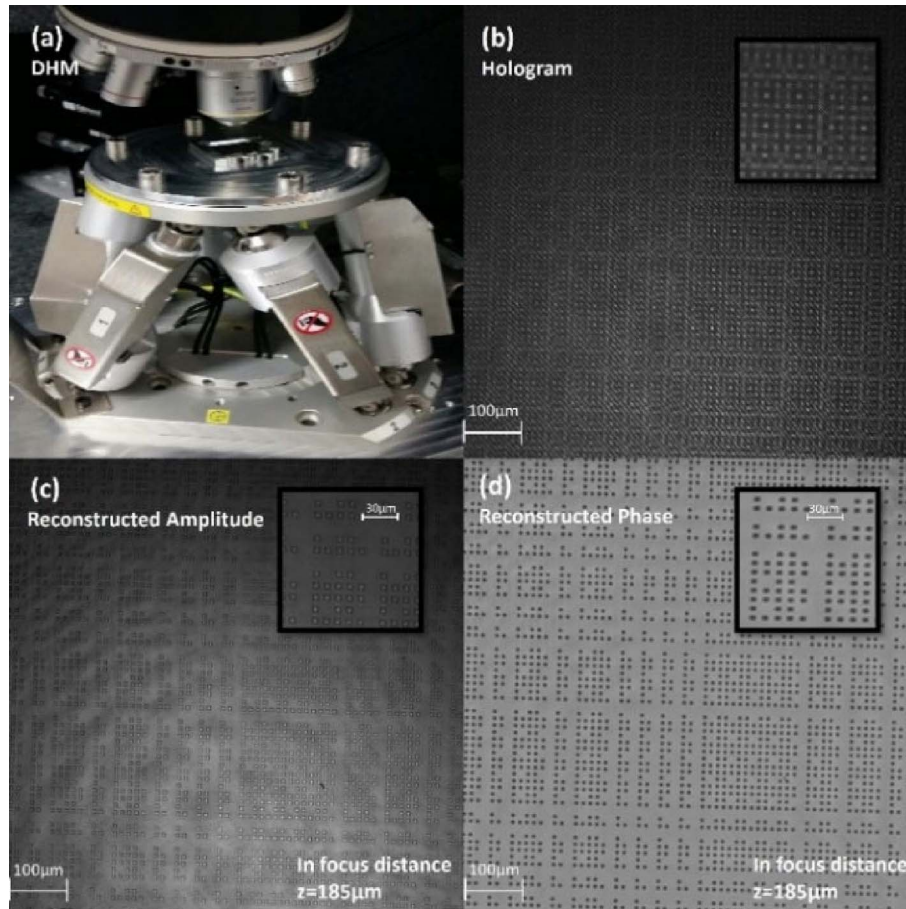
## 2 Theoretical background and context

### 2.1 Deep neural networks

DNNs inspired by biological neural networks, process, classify, and predict complex data through multi-layer structures. These networks employ non-linear transformations from input to output layers, enabling tasks like linearization in higher-dimensional spaces [4]. Optimization of DNN results involves a learning step, training the network with input-output data pairs. Adequate training data volume is crucial for optimal performance. DNNs, notably convolutional CNNs and ViT models, have demonstrated high effectiveness in tasks like image classification, computer vision, and solving complex problems such as autofocusing in DH [2, 3].

### 2.2 Digital holographic microscopy and computer micro-vision for micro-robotics

DH is an advanced imaging technique capturing both amplitude and phase of an object's entire wavefield using a CMOS imaging sensor. In Figure 1, we show typical experimental digital hologram a 2D pseudo-periodic pattern



**Figure 1.** (a) Lyncee-tec DHM observing a micro-structured pattern moved by a hexapod stage. (b) A typical experimental hologram of a pseudo-periodic pattern that allow 3D pose measurement [2]. Image reconstruction (c) in amplitude and (d) in phase at a numerical in-focus distance of 185  $\mu\text{m}$ .

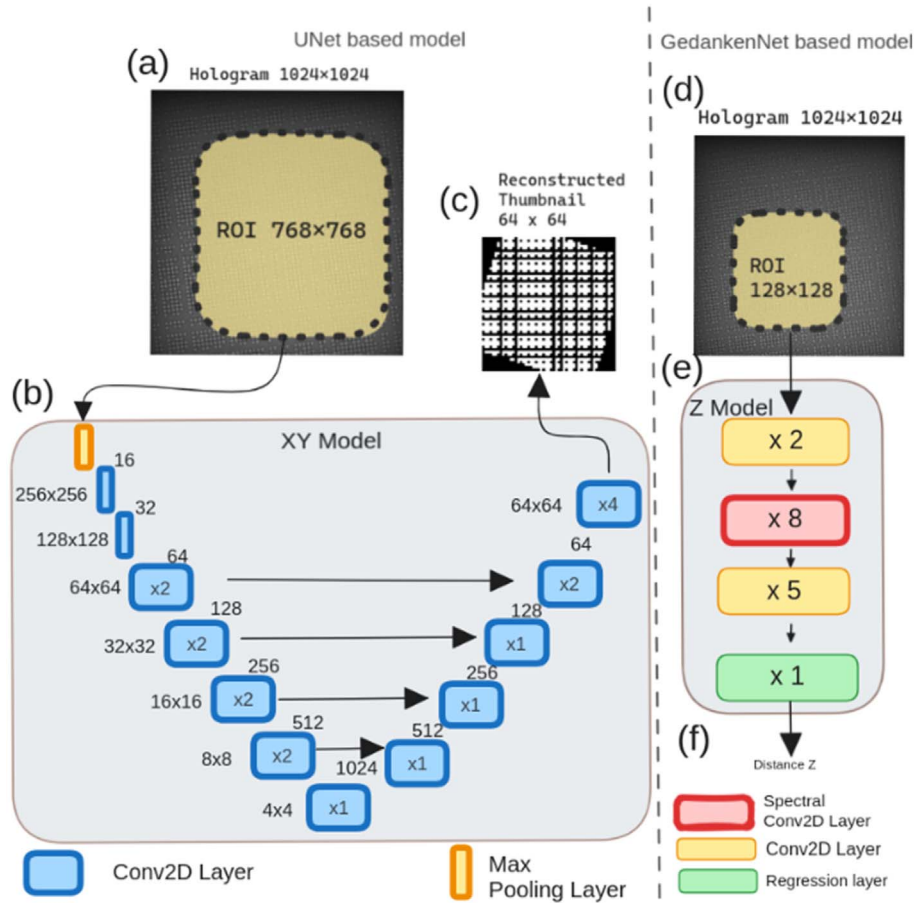
as phase object to perform 3D pose control in 3D through a microscope [2]. This study explores DH coupled with a computer micro-vision approach, employing phase correlation image processing techniques for sub-voxel sample pose measurements in micro-robotics [6, 7].

Digital hologram reconstruction relies on the Angular Spectrum Method [8], and a Lyncee-Tec Digital Holographic Microscope (DHM) equipped with  $10\times$  MO lens, adapts these principles to micro-objects, see reference [2] for experimental details. DHM works with digital autofocusing, enables automated microscopy and 3D pose control of micro-objects. Recent research highlights the use of DNN for faster auto-autofocusing in DHM through statistical image reconstruction, treating autofocusing as a classification or regression task [5]. The challenges include improving multiscale sensitivity for automated microscopy in 6 degrees of freedom (DoF) pose estimation while maintaining a broad field of view and depth of field [1]. A 2D pseudo-periodic pattern serves as a referencing sample (Fig. 1(c) and (d)). High-tech micro-assembly platforms in robotics demand translation and rotation stages (Fig. 1(a)), addressing increasingly complex tasks with nanoscale positioning resolution and large-scale movements beyond the centimetre range. This work addresses the challenge to target 3D

inference and video-rate control of samples for complex micro-nano manipulation such as 3D MEMS micro-nano-assembly and alignment, 3D nanoprinting, visual servoing for 3D nanopositioning [1].

### 3 Positioning models (X, Y and Z)

In this work, we combine previous autofocusing with DHM accelerated with DNN [2] giving Z position and a new approach to determine in the same time X and Y coordinates. In Figure 2(a–c), the structure of the XY Model (consisting of a series of 2D Convolution Layers and Max Pooling Layers) based on the UNet architecture [4] is presented, specifically designed for 3D pose estimation. The model takes a Region of Interest (ROI) extracted from the input hologram, initially sized at  $768 \times 768$  pixels within a hologram of  $1024 \times 1024$  pixels. The resulting output from the model is a reconstructed thumbnail of  $64 \times 64$ -pixels, encapsulating the X and Y positional information [6]. Subsequently, Figure 2(d–f) outlines the arrangement of the Z Model, which is based on an adapted version of a GedankenNet model proposed in [3]. The primary distinctions from the original version are that it



**Figure 2.** (a–c) Thumbnail reconstruction. (d–f) Assess the distance Z. (a) A ROI of  $768 \times 768$  is cropped from the hologram at a fixed position. (b) XY Model (based on a UNet like model). (c) The reconstructed thumbnail of  $64 \times 64$  pixels. (d) A ROI of  $128 \times 128$  is randomly cropped from the hologram space. (e) Z model based on an adapted version of a GedankenNet model [3]. (f) The distance Z.

accepts a single image as input and the input size has been minimized to  $128 \times 128$  pixels for faster computation of the Spectral Conv2D Layers (Fig. 2(f)). The XY Model's uniqueness lies in not reconstructing an image of the same size as the input (Fig. 2(b) depicts the initial Conv2D layers downsizing the input to  $64 \times 64$ ).

## 4 Methodology

We address this issue by applying DNNs to micro-vision measurement of 3D trajectories with DH. Recently, we demonstrated the ability of new generation of deep neural networks such as ViT to predict the in-focus distance with a high accuracy [2]. In a previous work, we also showed the ability of 2D pseudo-periodic pattern combined to conventional imaging system, used as in-plane position encoder, has allowed a 108 range-to resolution ratio through robust phase-based decoding [7]. Here, we present DNNs dedicated to hybrid approach combining computer micro-vision and DHM, able to perform simultaneously in-plane and out-plane measurements, at video-rate and without in focus full image reconstruction. The experimental setup is presented

in Figure 1. It consists in a DHM, a hexapod capable of precise motions along the 6DoF and a micro encoded pattern. We also show a typical hologram obtained and its reconstruction (Fig. 1(b)). The interferometric character of DH converts out-of-plane position of the sample in phase data that, combined with in-plane information retrieved from the micro-structured pattern, allows accurate measurement of 3D trajectories. DNNs speed up data processing and infer video-rate position detection.

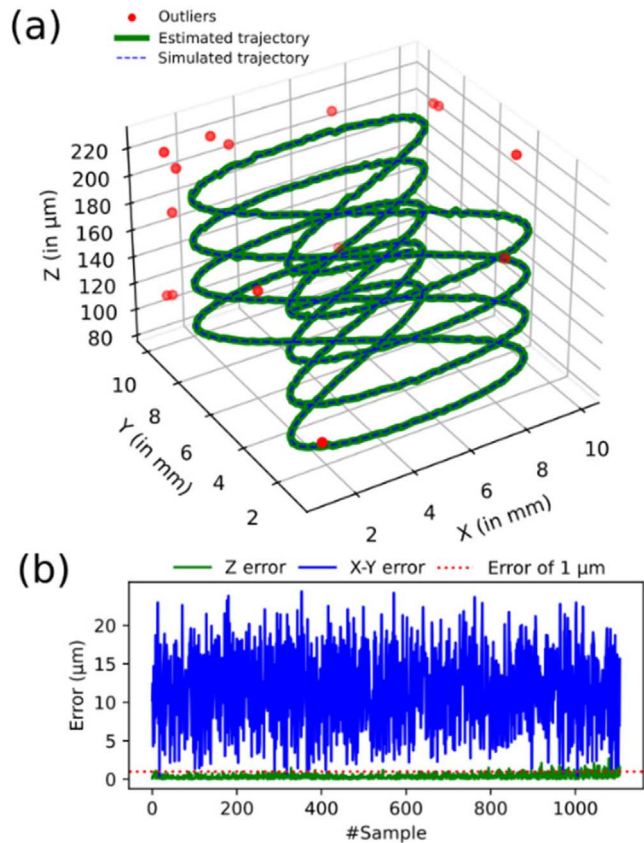
DNNs require training to realize expected tasks and to reach the best performances. In our work, the training step is conducted from a dataset constituted by simulated holograms. Various experimental parameters have been considered in simulations such as spherical aberration introduced by objective microscope lens, and has been implemented in simulated hologram datasets, with the aim of being able to mimic real experimental conditions. To rigorously evaluate the effectiveness of the proposed methodology, which integrates DH with DNNs and video-rate micro-vision, we conducted a comprehensive validation through simulation. Our primary objective was to assess the DNNs capability to predict a simulated 3D trajectory under precisely controlled conditions. For this purpose, we selected a Lissajous' figure

(result of superposing two harmonic motions on the X-Y plane). This complex trajectory served as a challenging yet well-defined path for rigorously testing the capabilities of the DH-DNN system. We simulated a complete 3D trajectory of 2D pseudo-periodic pattern with period of  $9\ \mu\text{m}$ , displaced by the hexapod stage (Fig. 1(a)), along the two-dimensional Lissajous trajectory in the X-Y plane and generated corresponding sequence of digital holograms. This trajectory was then extended into the third dimension by introducing incremental steps along the Z-axis, simulating motion in depth. Each step in the Z-direction corresponds to a subsequent holographic reconstruction distance for the simulated hologram. Subsequently, the generated holographic datasets were used in DNNs for training step and infer the trajectory. The networks were tasked with accurately predicting the Lissajous' trajectory based on the holographic dataset inputs, essentially capturing and replicating the complex curve in their predictions. To analyse each hologram (inference mode), both models are used (Fig. 2), XY Model and Z Model to get the associated thumbnail and Z distance. A post-processing algorithm is applied on the reconstructed thumbnail to extract the binary vectors representing the positions (X and Y) (Fig. 2c). To convert the binary vectors into meaningful micron-scale coordinates, each vector within the complete sequence of bits is identified. Those indexes are used to compute the final X and Y coordinates as described in [6].

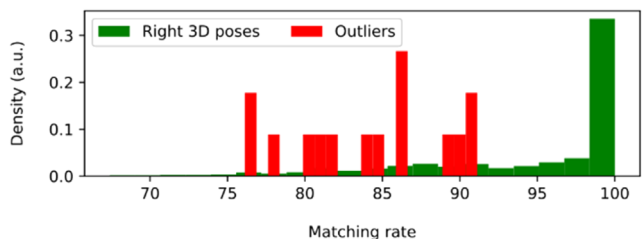
## 5 Results

We present the results obtained from the DH-DNN system methodology for predicting 3D trajectories. The models (XY Model and Z Model) have been trained using a total of 65000 simulated holograms. The XY Model is using binary cross entropy loss. The Z Model has been trained using a cross-validation method using the TanhExp loss function [9]. Both models are trained using the Adam optimizer. The models have been tested on a simulated trajectory of 1121 holograms. In Figure 3(a), the list of outliers (red points), the simulated (dashed blue line) and estimated (green line) trajectories are shown in 3D space. The accuracy exceeds 98% which demonstrates the system's ability to correctly estimate the 3D poses. Figure 3(b) provides a visual representation of the error along the Z axis and the deviation on the X-Y plane (L2-norm). This graphical depicts the precision of DNN predictions, revealing a max error of  $25\ \mu\text{m}$  on X-Y and less than  $1\ \mu\text{m}$  on Z. This X, Y level of performance must be compared with a maximum encoded area of  $11 \times 11\ \text{cm}^2$ . This allows video-rate monitoring of large displacements with a coarse but sufficient accuracy whereas eventual fine 3D pose is controlled by highly accurate but much slower conventional processing.

Figure 4 shows the matching rate associated to each estimated 3D pose. This underscores that a rate level between 90 and 100 is adequate for accurately decoding the correct position. The precision along the Z axis is of the same magnitude as in [2]. These results emphasize the DH-DNN methodology's capability to provide highly accurate and detailed predictions of three-dimensional



**Figure 3.** (a) Outliers (in red), simulated (in blue) and estimated (in green) trajectory in the 3D space. (b) Z and X-Y errors in  $\mu\text{m}$  (absolute difference and L2-norm). The Z error is mostly below an error of  $1\ \mu\text{m}$  (red dashed line).



**Figure 4.** Matching rate associated to each 3D pose (red: outliers, green: right 3D poses).

trajectories. This highlights its practical utility in real-time micro-robotics and micro-vision applications. Moreover, the average inference speed is below 20 ms on a NVidia RTX 3090 32 GB mainly consumed by the data transfer of the images to the GPU (XY Model: 7.5 ms inference; Z Model: 2.5 ms inference; 10 ms for the data transfer).

## 6 Conclusions

We propose a method that enables the direct determination of 3D positions from hologram space with a mean error of

1  $\mu\text{m}$  on Z and 12  $\mu\text{m}$  on X-Y, effectively bypassing the need for full holographic image reconstruction. These errors must be compared to the complete encoded area of  $11 \times 11 \text{ cm}^2$ . Moreover, our study offers a thorough analysis of the matching rate levels attributed to each 3D pose. We believe it is the first time a GedankenNet model is used as a regression tool. The modified GedankenNet (Z Model) achieved an inference speed of 2.5 ms, contrasting with the over 20 ms required by a TViT [2].

### Funding

This work was supported by Agence Nationale de la Recherche, HOLO-CONTROL project (Contract No. ANR-21-CE42-0009), by the French Investissements d'Avenir program, TIRREX project (Contract No. ANR-21-ESRE0015), SMARTLIGHT project (ANR-21-ESRE-0040) and by Cross-disciplinary Research (EIPHI) Graduate School (contract No. ANR-17-EURE-0002). This work was performed using HPC resources from GENCIIDRIS (Grant 20XX-AD011012913R2) and the Mésocentre de Franche-Comté.

### Conflicts of interest

The authors declare that they have no competing interests to report.

### Data availability statement

The data associated with this study is available upon request. Please contact the corresponding author to request access to the data.

### Author contribution statement

PS, RC, GJL and MJ contributed to the conceptualization of the idea. The development of deep neural networks and datasets was performed by SC. The experiments and simulations of data were performed by BA and JBC. The development of digital hologram algorithms was performed by JBC. SC, JBC and MJ wrote the manuscript with feedback from RC, GJL and PS. All authors discussed the results and contributed to the final manuscript.

### References

- 1 Yao S., Li H., Pang S., Zhu B., Zhang X., Fatikow S. (2021) *IEEE Trans. Instrum. Meas.* **70**, 1–28.
- 2 Cuenat S., Andréoli L., André A.N., Sandoz P., Laurent G.J., Couturier R., Jacquot M. (2022) *Opt. Express* **30**, 14.
- 3 Huang L., Chen H., Liu T., et al. (2023) Self-supervised learning of hologram reconstruction using physics consistency, *Nat. Mach. Intell.* **5**, 895–907.
- 4 Ronneberger O., Fischer P., Brox T. (2015) arXiv. 1505.04597.
- 5 Zeng T., Zhu Y., Lam E.Y. (2021) *Opt. Express* **29**, 24.
- 6 André A.N., Sandoz P., Mauzé B., Jacquot M., Laurent G.J. (2022) *Int. J. Comput. Vis.* **130**, 6.
- 7 André A.N., Sandoz P., Mauzé B., Jacquot M., Laurent G.J. (2020) *EEE/ASME Trans. Mech.* **25**, 1193–1201.
- 8 Goodman J.W. (2005) *Introduction to fourier optics*, Roberts & Company Publishers, Englewood, pp. 55–61.
- 9 Liu X., Di X. (2020) arXiv. 2003.09855.