

1 Digital holographic microscopy applied to 3D computer micro-vision 2 by using deep neural networks

3 *Stéphane Cuenat*^{1,2}, *Jesús E. Brito Carcaño*^{1,2}, *Belal Ahmad*¹, *Patrick Sandoz*¹, *Raphaël Couturier*¹, *Guillaume*
4 *Laurent*¹ and *Maxime Jacquot*^{1,*}

5 ¹Université de Franche-Comté, SUPMICROTECH-ENSMM, CNRS, Institut FEMTO-ST, 25000 Besançon, France

6 ²Jesús E. Brito Carcaño and Stéphane Cuenat contributed equally to this work as first authors

7 **Abstract.** Deep neural networks (DNNs) are increasingly employed across diverse fields
8 of applied science, particularly in areas like computer vision and image processing, where
9 they enhance the performance of instruments. Various advanced coherent imaging
10 techniques, including digital holography, leverage different deep architectures like
11 convolutional neural networks (CNN) or Vision Transformers (ViT). These architectures
12 enable the extraction of diverse metrics such as autofocusing reconstruction distance or 3D
13 position determination, facilitating applications in automated microscopy and phase image
14 restitution. In this work, we propose a hybrid approach utilizing an adapted version of the
15 GedankenNet model, coupled with a UNet-like model, for the purpose of accessing micro-
16 objects 3D pose measurements. These networks are trained on simulated holographic
17 datasets. Our approach achieves an accuracy of 98% in inferring the 3D poses. We show
18 that a GedankenNet can be used as a regression tool and is faster than a Tiny-ViT (TViT)
19 model. Overall, integrating deep neural networks into digital holographic microscopy and
20 3D computer micro-vision holds the promise of significantly enhancing the robustness and
21 processing speed of holograms for precise 3D position inference and control, particularly
22 in micro-robotics applications.

23 1 Introduction

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

32 2. Theoretical Background and context

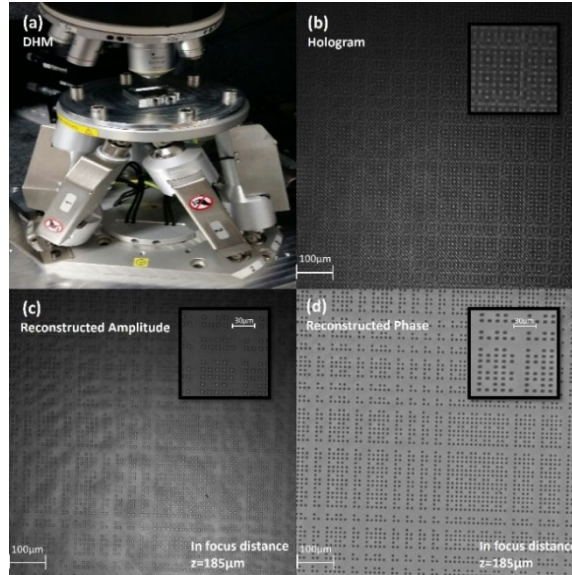
33 2.1 Deep Neural Networks

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

*Corresponding author: maxime.jacquot@univ-fcomte.fr

40 2.2 Digital holographic microscopy and computer micro-vision for micro-robotics

41 DH is an advanced imaging technique capturing both amplitude and phase of an object's entire wavefield using
42 a CMOS imaging sensor. In Fig. 1, we show typical experimental digital hologram a 2D pseudo-periodic
43 pattern as phase object to perform 3D pose control in 3D through a microscope [2]. This study explores DH
44 coupled with a computer micro-vision approach, employing phase correlation image processing techniques for
45 sub-voxel sample pose measurements in micro-robotics [6, 7].



46

47 **Fig. 1.** (a) Lyncee-tec DHM observing a micro-structured pattern moved by a hexapod stage . (b) A typical
48 experimental hologram of a pseudo-periodic pattern that allow 3D pose measurement [2]. Image reconstruction (c) in
49 amplitude and (d) in phase at a numerical in-focus distance of 185µm.

50 Digital hologram reconstruction relies on the Angular Spectrum Method [8], and a Lyncee-Tec Digital
51 Holographic Microscope (DHM) equipped with 10x MO lens, adapts these principles to micro-objects, see
52 reference [2] for experimental details. DHM works with digital autofocusing, enables automated microscopy
53 and 3D pose control of micro-objects. Recent research highlights the use of DNN for faster auto-autofocusing
54 in DHM through statistical image reconstruction, treating autofocusing as a classification or regression task
55 [5]. The challenges include improving multiscale sensitivity for automated microscopy in 6 degrees of
56 freedom (DoF) pose estimation while maintaining a broad field of view and depth of field [1]. A 2D pseudo-
57 periodic pattern serves as a referencing sample (Fig. 1(c) and (d)). High-tech micro-assembly platforms in
58 robotics demand translation and rotation stages (Fig. 1(a)), addressing increasingly complex tasks with
59 nanoscale positioning resolution and large-scale movements beyond the centimetre range. This work addresses
60 the challenge to target 3D inference and video-rate control of samples for complex micro-nano manipulation
61 such as 3D MEMS micro-nano-assembly and alignment, 3D nanoprinting, visual servoing for 3D
62 nanopositioning [1].

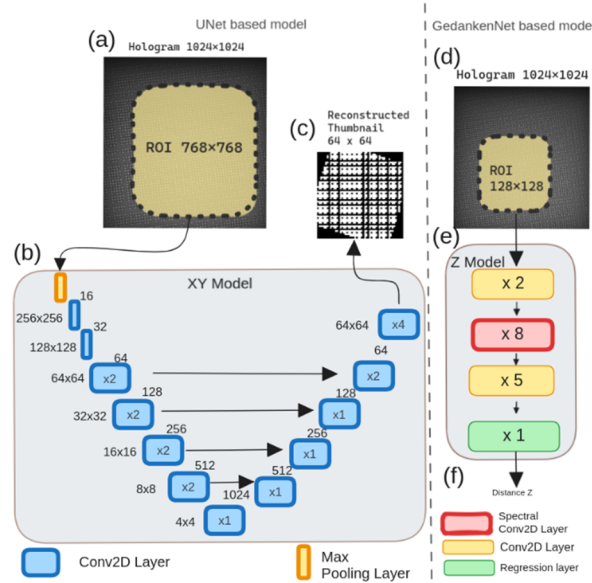
63 3 Positioning Models (X, Y and Z)

64 In this work, we combine previous autofocusing with DHM accelerated with DNN [2] giving Z position and
65 a new approach to determine in the same time X and Y coordinates. In Fig. 2(a-c), the structure of the XY
66 Model (consisting of a series of 2D Convolution Layers and Max Pooling Layers) based on the UNet
67 architecture [4] is presented, specifically designed for 3D pose estimation. The model takes a Region of Interest
68 (ROI) extracted from the input hologram, initially sized at 768x768 pixels within a hologram of 1024x1024
69 pixels. The resulting output from the model is a reconstructed thumbnail of 64x64-pixels, encapsulating the X

70 and Y positional information [6]. Subsequently, Fig. 2(d-f) outlines the arrangement of the Z Model, which is
 71 based on an adapted version of a GedankenNet model proposed in [3]. The primary distinctions from the
 72 original version are that it accepts a single image as input and the input size has been minimized to 128x128
 73 pixels for faster computation of the Spectral Conv2D Layers (Fig. 2(f)). The XY Model's uniqueness lies in
 74 not reconstructing an image of the same size as the input (Fig. 2(b) depicts the initial Conv2D layers
 75 downsizing the input to 64x64).

76 4 Methodology

77 We address this issue by applying DNNs to micro-vision measurement of 3D trajectories with DH. Recently,
 78 we demonstrated the ability of new generation of deep neural networks such as ViT to predict the in-focus
 79 distance with a high accuracy [2]. In a previous work, we also showed the ability of 2D pseudo-periodic
 80 pattern combined to conventional imaging system, used as in-plane position encoder, has allowed a 108
 81 range-to resolution ratio through robust phase-based decoding [7]. Here, we present DNNs dedicated to
 82 hybrid approach combining computer micro-vision and DHM, able to perform simultaneously in-plane and
 83 out-plane measurements, at video-rate and without in focus full image reconstruction. The experimental
 84 setup is presented in Fig. 1. It consists in a DHM, a hexapod capable of precise motions along the 6DoF and
 85 a micro encoded pattern. We also show a typical hologram obtained and its reconstruction (Fig. 1 (b)). The
 86 interferometric character of DH converts out-of-plane position of the sample in phase data that, combined
 87 with in-plane information retrieved from the micro-structured pattern, allows accurate measurement of 3D
 88 trajectories. DNNs speed up data processing and infer video-rate position detection.



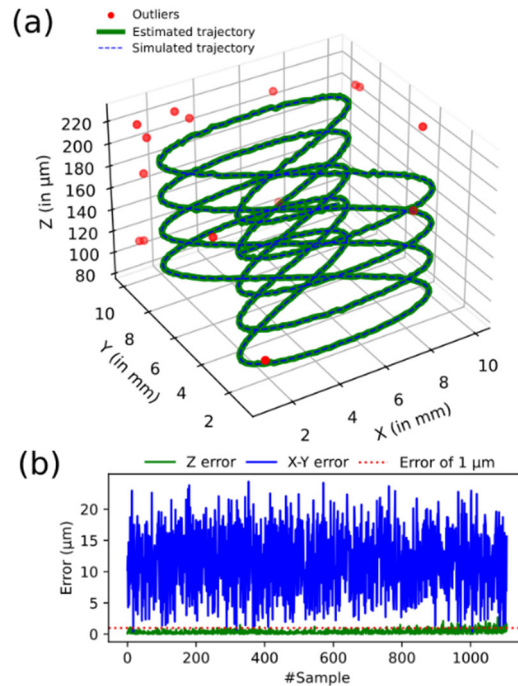
89
 90
 91 **Fig. 2.** (a-c) Thumbnail reconstruction. (d-f) Assess the distance Z. (a) A ROI of 768x768 is cropped from the hologram
 92 at a fixed position. (b) XY Model (based on a UNet like model). (c) The reconstructed thumbnail of 64x64 pixels. (d) A
 93 ROI of 128x128 is randomly cropped from the hologram space. (e) Z model based on an adapted version of a
 94 GedankenNet model [3]. (f) The distance Z.

95 DNNs require training to realize expected tasks and to reach the best performances. In our work, the training
 96 step is conducted from a dataset constituted by simulated holograms. Various experimental parameters have
 97 been considered in simulations such as spherical aberration introduced by objective microscope lens, and has
 98 been implemented in simulated hologram datasets, with the aim of being able to mimic real experimental
 99 conditions. To rigorously evaluate the effectiveness of the proposed methodology, which integrates DH with
 100 DNNs and video-rate micro-vision, we conducted a comprehensive validation through simulation. Our primary
 101 objective was to assess the DNNs capability to predict a simulated 3D trajectory under precisely controlled
 102 conditions. For this purpose, we selected a Lissajous' figure (result of superposing two harmonic motions on
 103 the X-Y plane). This complex trajectory served as a challenging yet well-defined path for rigorously testing

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

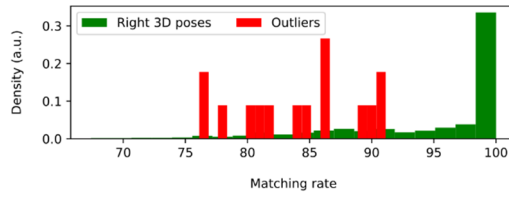
117 5 Results

118 We present the results obtained from the DH-DNN system methodology for predicting 3D trajectories. The
 119 models (XY Model and Z Model) have been trained using a total of 65000 simulated holograms. The XY
 120 Model is using binary cross entropy loss. The Z Model has been trained using a cross-validation method using
 121 the TanhExp loss function [9]. Both models are trained using the Adam optimizer. The models have been tested
 122 on a simulated trajectory of 1,121 holograms. In Fig. 3(a), the list of outliers (red points), the simulated (dashed
 123 blue line) and estimated (green line) trajectories are shown in 3D space. The accuracy exceeds 98% which
 124 demonstrates the system's ability to correctly estimate the 3D poses. Fig. 3(b) provides a visual representation
 125 of the error along the Z axis and the deviation on the X-Y plane (L2-norm). This graphical depicts the precision
 126 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
 127 performance must be compared with a maximum encoded area of $11 \times 11\ \text{cm}^2$. This allows video-rate monitoring
 128 of large displacements with a coarse but sufficient accuracy whereas eventual fine 3D pose is controlled by
 129 high accurate but much slower conventional processing.



130

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



133
134 **Fig. 4.** Matching rate associated to each 3D pose (red: outliers, green: right 3D poses).

135 Fig. 4 shows the matching rate associated to each estimated 3D pose. This underscores that a rate level between
136 90 and 100 is adequate for accurately decoding the correct position. The precision along the Z axis is of the
137 same magnitude as in [2]. These results emphasize the DH-DNN methodology's capability to provide highly
138 accurate and detailed predictions of three-dimensional trajectories. This highlights its practical utility in real-
139 time micro-robotics and micro-vision applications. Moreover, the average inference speed is below 20 ms on
140 a NVidia RTX 3090 32 GB mainly consumed by the data transfer of the images to the GPU (XY Model: 7.5
141 ms inference; Z Model: 2.5 ms inference; 10 ms for the data transfer).

142 6 Conclusions

143 We propose a method that enables the direct determination of 3D positions from hologram space with a mean
144 error of 1 μm on Z and 12 μm on X-Y, effectively bypassing the need for full holographic image reconstruction.
145 These errors must be compared to the complete encoded area of 11x11cm². Moreover, our study offers a
146 thorough analysis of the matching rate levels attributed to each 3D pose. We believe it is the first time a
147 GedankenNet model is used as a regression tool. The modified GedankenNet (Z Model) achieved an inference
148 speed of 2.5 ms, contrasting with the over 20 ms required by a TViT [2].

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

155 References

- 156 1. S. Yao, H. Li, S. Pang, B. Zhu, X. Zhang, S. Fatikow IEEE Trans. Instrum. Meas, **70**, (2021)
157 2. S. Cuenat, L. Andréoli, A.N. André, P. Sandoz, G.J. Laurent, R. Couturier, M. Jacquot, Opt. Express. **30**, 14 (2022)
158 3. Huang, L., Chen, H., Liu, T. et al. Self-supervised learning of hologram reconstruction using physics consistency.
159 Nat Mach Intell 5, 895–907 (2023).
160 4. Olaf Ronneberger and Philipp Fischer and Thomas Brox, arXiv. 1505.04597 (2015)
161 5. T. Zeng, Y. Zhu, E. Y. Lam, Opt. Express **29**, 24 (2021)
162 6. A.N. André, P. Sandoz, M. Jacquot, G.J. Laurent, Int. J. Comput. Vis. **130**, 6 (222)
163 7. A.N. André, P. Sandoz, B. Mauzé, M. Jacquot, G. J. Laurent, EEE/ASME Trans. on Mech. **25** (2020)
164 8. Goodman J.W. (2005) *Introduction to Fourier Optics*, Roberts & Company Publishers, pp. 55-61.
165 9. Xinyu Liu and Xiaoguang Di, arXiv. 2003.09855 (2020)