Nouvelles architectures de réseaux de neurones profonds pour un autofocus rapide en microscopie holographique numérique

Louis Andréoli, Stéphane Cuenat, Jesus E. Brito Carcaño, Antoine N. André, Patrick Sandoz, Raphaël Couturier, Guillaume J. Laurent, Maxime Jacquot

Institut FEMTO-ST, CNRS & Université Bourgogne Franche-Comté
15B avenue des Montboucons, 25000 Besançon, France

maxime.jacquot@univ-fcomte.fr

RESUME

La détermination précise de la position axiale d'un objet sur une large étendue reste un véritable enjeu. En réalisant l'hologramme expérimental d'une cible avec un objectif de microscope x10 et en entrainant des réseaux profonds type *Transformer*, la distance est inférée en 25 ms sur CPU avec une précision de 1.2 µm sur une plage totale de 92 µm.

MOTS-CLEFS : *HOLOGRAPHIE, AUTOFOCUS, MICROSCOPIE, RESEAUX PROFONDS TRANSFORMER*

1. INTRODUCTION

L'intérêt récent pour le Deep Learning (DL) s'est étendu à divers domaines scientifiques, allant des neurosciences à l'informatique [1] en passant par la physique [2]. Les enjeux liés aux techniques d'imagerie avancées, tels que la microscopie pour le micro-positionnement 3D ou le suivi en temps réel, sont aussi largement étudiés dans une perspective d'apprentissage automatique. En effet, la microscopie optique fournit un outil pratique sans contact pour les mesures de position et de déplacement d'objets en 3D. Cependant, la courte profondeur de champ des objectifs de microscope limite considérablement les déplacements autorisés le long de la direction axiale. En outre, la qualité de l'image et le contenu de l'information ne dépendent pas uniquement des instruments utilisés.

2. NOUVELLES TENDANCES DE L'APPRENTISSAGE PROFOND POUR LE TRAITEMENT D'IMAGES EN HOLOGRAPHIE

Les approches d'imagerie computationnelle peuvent encore être optimisées avec des algorithmes de DL. L'holographie numérique (HN) permet de reconstruire les informations optiques sur de grandes plages axiales grâce au calcul de netteté de l'image. Cette approche est alors très efficace pour déterminer le positionnement 3D d'un échantillon, sans mouvement mécanique. La mise au point automatique est un problème difficile qui peut être étudié comme une tâche de régression, où les paramètres physiques d'imagerie ne sont plus nécessaires. Ce changement de paradigme permet une localisation précise de l'échantillon en profondeur sans reconstructions d'images, permettant ainsi une mesure de positionnement en temps réel. Nos travaux explorent les capacités visuelles étendues offertes par la combinaison d'algorithmes HN et DL de dernière génération telle que le réseau *Vision Transformer* (ViT) et *Swin-Transformer* (SwinT) pour des applications en microrobotique [3] ou en 3D temps réel microscopie [4]. Depuis 2020, les algorithmes *Transformer*, initialement créés pour le traitement du langage naturel, sont de plus en plus utilisés pour les tâches de classification d'images [5] et peuvent atteindre des performances supérieures sur les tâches de classification d'images [6]. C'est la première fois, à notre connaissance, que des algorithmes *Transformer* sont appliqués à l'imagerie cohérente telle que l'holographie.

3. RESULTATS

À l'aide d'un microscope holographique numérique (DHM de Lyncee Tec Corp, MO 10x, NA 0,3, λ = 675 nm), nous avons enregistré 40 000 hologrammes numériques d'une cible pseudo-périodique [7] se déplaçant en 3D sur une plage axiale de 92 µm. Par rapport aux architectures standards, nous l'avons simplifiée les réseaux proposés (TViT, TSwinT, TVGG) pour obtenir des versions réduites possédant un nombre d'hyperparamètres 10 à 40 fois inférieur. Après l'apprentissage des réseaux avec les données expérimentales, la position axiale de l'échantillon sur toute la plage considérée est déterminée avec une précision micrométrique (cf Fig.1(a)), c'est-à-dire dépassant d'un facteur 12 la limite de résolution axiale de l'objectif du microscope. Nous avons également testé les réseaux avec des hologrammes simulés sans bruit optique et nous avons déterminé une première limite de résolution axiale inférieure à 0,3 µm (cf Fig.1(b)). Utilisant un CPU, le temps d'inférence des trois modèles est inférieur à 50 ms sur Intel i7 et inférieur à 25 ms sur Intel i9, ce qui est entièrement compatible avec des applications en temps réel, sans forcément l'utilisation de cartes GPU couteuses.

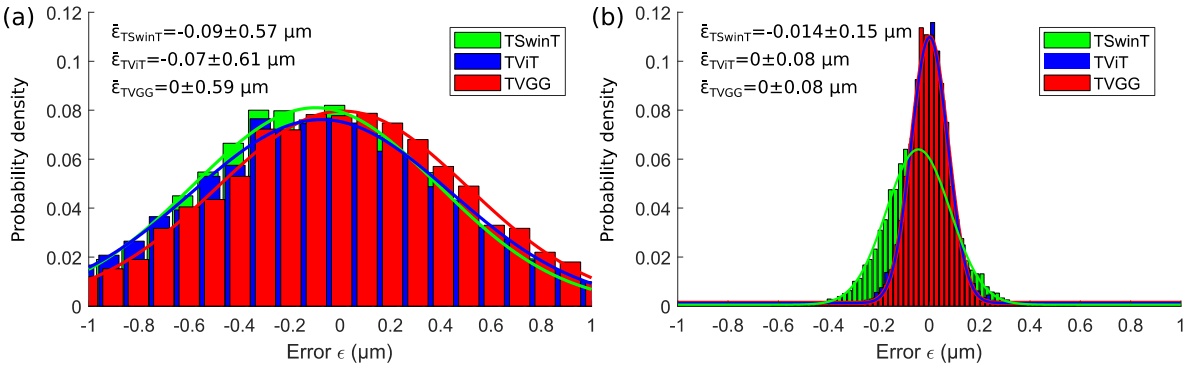

Figure 1. Distribution des erreurs de distance de reconstruction pour les trois réseaux de neurones. (a) et (b) sont respectivement les résultats pour les modèles expérimentaux et simulés.

RÉFÉRENCES

[1]. Azar, J., Makhoul, A., Barhamgi, M., & Couturier, R. An energy efficient IoT data compression approach for edge machine learning. *Future Generation Computer Systems*, *96*, 168-175. (2019).

[2]. Larger L., Baylon Fuentes A., Martinenghi R., Udaltsov V., Chembo Kouomou Y. and Jacquot M. (2017). High-Speed Photonic Reservoir Computing Using a Time-Delay-Based Architecture: Million Words per Second Classification. *Physical Review X, vol. 7,* pp. 011015-1/14.

[3]. André, A. N., Sandoz, P., Mauzé, B., Jacquot, M., & Laurent, G. J. (2020). Sensing one nanometer over ten centimeters: A microencoded target for visual in-plane position measurement. *IEEE/ASME Transactions on Mechatronics,* 25(3), 1193-1201.

[4]. H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, "Deep learning for single-shot autofocus microscopy," Optica 6, 794–797 (2019).

[5]. Dosovitskiy, A., Beyer, L., Kolesnikov, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. *Preprint arXiv:2010.11929*, 2020.

[6]. M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, and A. Dosovitskiy, "Do vision transformers see like convolutional neural networks?" (2021).

[7]. Sandoz, P., and Jacquot, M. (2011). Lensless vision system for in-plane positioning of a patterned plate with subpixel resolution. *JOSA A,* 28(12), 2494-2500.