Using Deep Learning for Object Distance Prediction in Digital Holography

Raphaël Couturier, Michel Salomon
FEMTO-ST Institute, CNRS
Univ. Bourgogne Franche-Comté (UBFC)
Belfort, France
{raphael.couturier,michel.salomon}@univ-fcomte.fr

Elie Abou Zeid, Chady Abou Jaoudé
Ticket Laboratory
Antonine University
Hadat-Baabda, Lebanon
{elie.abouzeid,chady.aboujaoude}@ua.edu.lb

***Abstract*—Deep Learning (DL) has marked the beginning of a new era in computer science, particularly in Machine Learning (ML). Nowadays, there are many fields where DL is applied such as speech recognition, automatic navigation systems, image processing, etc [1]. In this paper, a Convolutional Neural Network (CNN), more precisely a CNN built on top of DenseNet169, is proven to be helpful in predicting object distance in computer-generated holographic images. The problem is addressed as a classification problem where 101 classes of images were generated, each class corresponding to a different distance value from the object at a micrometer scale. Experiments show that the proposed network is efficient in this context, being able to classify with a 100% accuracy level if trained properly.**

Keywords-deep learning; convolutional neural networks; holographic images; digital holography

I. INTRODUCTION

Digital holography (DH) is an emerging field in imaging applications [2,3]. In fact, it is exploited in 3D image processing, surface contour measurements, microscopy [4], and even in microrobotics [5]. A major challenge in microrobotics and/or photonics is to be able to determine metrics in the context of complex imaging devices. Among those metrics, it is interesting to investigate how the 3D position measurements of micro-objects can be determined. With earlier hologram image reconstruction, the overall experimental setup needed to be determined ahead of time including object's depth position [6], otherwise many diffraction calculations were needed to be applied for various depth settings. According to [7] and [8], such techniques consumed a lot of time as they necessitate many diffraction calculations and signal processing, and thus heavy computations were required in this scope. Nowadays, Deep Learning (DL) reshaped the world of computer science and it is used in many application areas including DH. Particularly, DL helped in coping the time consumption and heavy computation concerns of the older techniques to determine depth position: instead of applying many diffraction calculations, training a deep neural network is adopted to enable it to do the depth predictions itself [6,9]. Such predictions can result from treating the considered problem as either a classification problem [10,11,12] or a regression problem [6,9].

In this paper, given the scale of the available dataset, the problem of determining the distance of a view captured by a holographic camera is addressed by modeling it as a classification problem. A major challenge that is tackled in this context is to design a deep neural network that is able to give accurate predictions, while working with high precision distance values, such as on a micrometer scale, with a considerable amount of classes to perform the classifications. The outcomes proved that the built network is in fact capable of performing predictions with an accuracy level of 100% at a micrometer scale, using 101 classes for the classifications, where each class corresponds to a different distance from the object varying with a step of 100 micrometer.

The proposal is presented thereafter throughout the following sections. Section II presents some of the existing literature, more specifically on deep learning and its benefits for addressing classification problems in digital holography. The context of this study is highlighted in the following section. Section IV is dedicated to a detailed presentation of the proposed deep neural network and the experimental results. Finally, the conclusion aims to summarize the findings and the potential future works.

II. RELATED WORKS

First, we will present DL globally before digging deeper into its use in digital holography. A CNN is a well-known DL architecture that is widely used for analyzing and classifying images by extracting and learning features directly from them [13]. Many CNN models are available, each having its own particularity and advantages. DenseNet [14] (a densely connected network) is one of them, which is known for its significant results when compared to other models. It even necessitates fewer variables for training [15]. According to the authors of this CNN model [14], DenseNet networks do not encounter optimization issues even upon scaling to hundreds of layers. Their main motivation to create this model was to cope the vanishing of the input information that occurs at the output of the network after it has passed through many layers, as well as the vanishing of the gradient in the opposite direction.

In the following, we will discuss some works in the literature that tackle DH using DL. Currently, many studies in DH focus on reconstructing the target object's amplitude and phase once its distance is extracted; a process referred to

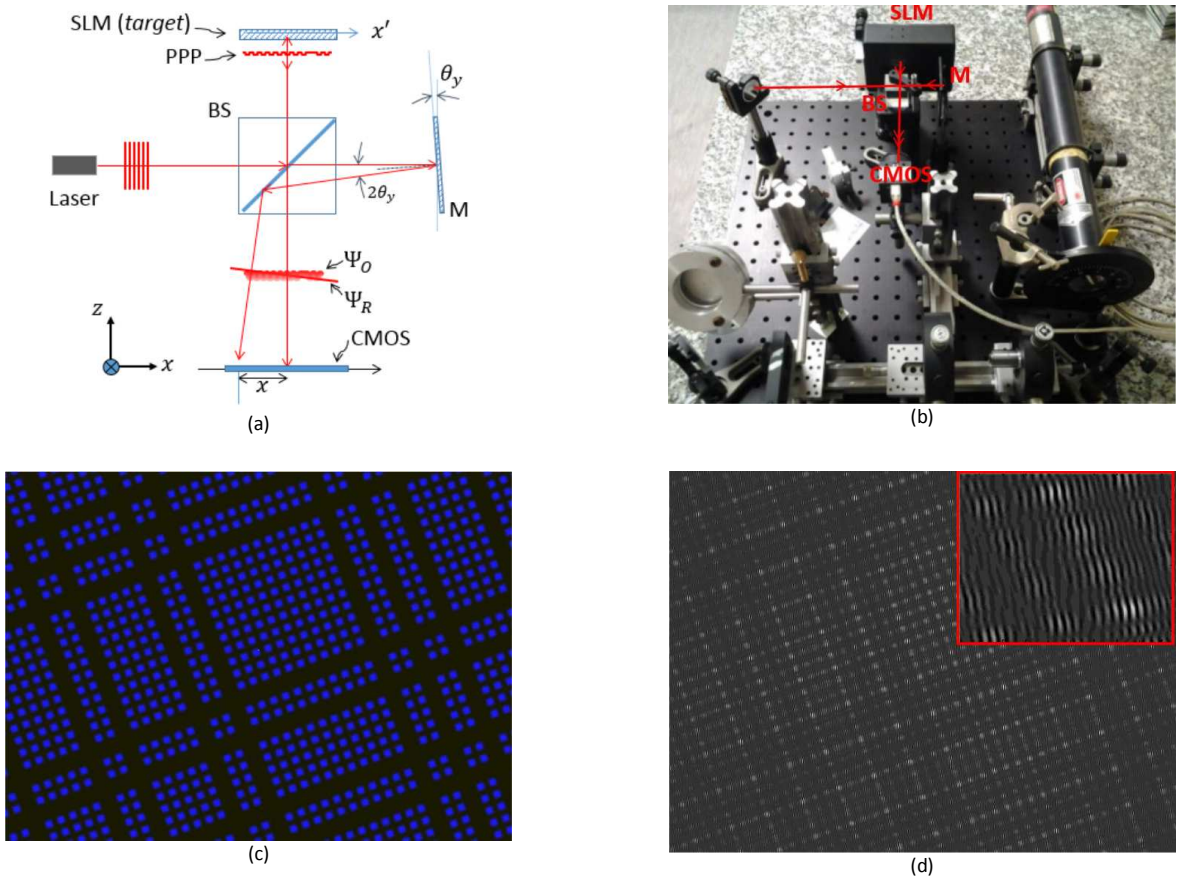

Figure 1. (a) Schematic drawing and (b) picture of the experimental setup, (c) pseudo-periodic pattern (total size 5.6 * 4.2 mm^2), (d) computed generated digital hologram – images drawn from [16].

as "autofocusing". In particular, identifying the object distance is crucial for the object reconstruction [12]. In the literature, autofocusing has been explored by exploiting advances in DL (also called learning-based approaches). To be more specific, two main approaches are investigated to tackle the problem: using classification [10,11,12] and using regression [6,9].

We will focus on the classification approach and its adoption in the literature. In this approach, [10] and [11] were the first to work on predicting depth in DH in discrete values. In [10], the authors demonstrated that autofocusing in DH can be achieved using DL. They adopted a CNN that is built on top of the AlexNet architecture. They started off with a number of holograms that they prepared at a first stage by eliminating the "zero-order" and the "twin-term". They used 21 classes for the labeling and data augmentation was applied on the prepared holograms to increase the size of the dataset. Moreover, the network was trained with manually defined depth values. 90% of the data were used to train the network and the remaining 10% for the validation. Their learning rate was set up to the value of 0.01. As they concluded, their network does not scale properly. Moreover, they stated that their network "generalizes well" at a millimeter scale without mentioning the obtained accuracy level to prove it.

In [12] the authors used a CNN to predict the object distance by training their network with hologram-specific labels that correspond to the actual distances. They used an uniform technical setup to capture 1000 holograms, ending up with 5 distance labels. Their experimental results proved that the network is able to predict the distance without any knowledge of the technical setup and with less time consumption than the traditional methods. This proves that they coped the issues that were encountered in the work [10]. However, although they provided estimates at a millimeter scale (axial range of about 3 mm, which is appropriate for imaging systems), they only worked on 5 classes, which may not be enough to generalize their findings.

In this paper, DenseNet [14] is used to build a CNN model that achieves depth prediction in the DH context. The difference between DenseNet and a conventional CNN is the use of additional connections between shallow layers and deeper ones. Indeed, in DenseNet each layer receives as inputs the feature maps from its preceding layers and sends its own maps to the deeper layers. Thus, each layer receives all the knowledge gathered by all the previous layers. In our case, DenseNet classifies in 101 classes and at a micrometer

scale, which is far more accurate than most of the existing works in literature, where efforts were channeled to work on a millimeter scale. It uses a dataset of 10,100 images that are augmented and split using the 70-30% rule, without using the training images in the testing. The model is pre-trained using Imagenet. Experimental results presented thereafter prove that the DenseNet network is able to predict depth with 100% accuracy.

III. Digital Holography for Measurements in Microrobotics

First the context of this study is presented. The goal is to explore the capabilities provided by digital holography (DH) coupled with digital photogrammetry to perform sub-pixel sample positioning measurements in microrobotics and in photonics. The determination of metrics in the context of complex imaging devices used in microrobotics and/or photonics represents a large challenge. At the end, the goal is to be able to perform micro-scale 3D position measurements of micro-objects. In this paper, a simpler goal is to be able to determine the distance of a sight (without object) taken by a holographic camera in order to simplify the problem. Because with deep learning, a large set of data is required and because physicists often build model of reality to better understand it, our physicist colleagues made a model of the holography camera.

As an illustration, Fig. 1 depicts images from a vision method based on digital holography developed by our physicist colleagues [16]. This method allows the simultaneous measurement of the in-plane position and orientation of a moving target object with sub-pixel resolution. Fig. 1 (a) and (b) show the experimental setup, where CMOS denotes the image sensor, BS the beam splitter, M the mirror, and PPP the pseudo-periodic pattern wavefront reflected from SLM/target device. Fig. 1 (c) and (d) presents, respectively, the image of the PPP obtained experimentally for a particular in-plane position and the corresponding hologram generated numerically on a computer. On the upper part of Fig 1 (d) is a zoom on a small part of the generated hologram showing the modulated fringe carrier that result from the tilt of the mirror M. Consequently, using a similar numerical model provided by our colleagues, we can generate as many images as required.

In order to understand the difficulty of the problem we are facing, in the following some images of sight are given (see Fig. 2). Distance of the sight may vary from 1 cm to 2 cm with steps of size 100 µm. So there are 100 different steps corresponding to 100 classes. It can be seen that sights can rotate freely and that there are sometimes holes in sights in order to let physicists code some information of objects directly visible with the DH. It should also be noticed that only a Region Of Interest (ROI) is presented and analyzed by the CNN in the following. In practice, a crop operation is performed on an image. In Fig. 2 it can be seen that images corresponding to neighboring depths are very similar. For a human, being able to recognize the distance is not possible.

IV. The Proposed Deep Network and Experimental Results

To tackle the classification problem, a good CNN network is needed. Therefore DenseNet was selected given that it has, according to its authors, significant results when compared to other models. Besides, many architecture variations of this model exist. At first, DenseNet-121 was adopted. However, the obtained results were not as accurate as expected and thus we decided to replace it with DenseNet-169. This latter DenseNet flavor effectively allowed having better results in the current context. Indeed, thanks to DenseNet-169 a free-error classification could be obtained, whereas the lighter DenseNet-121 always failed on a few dozens of images. In fact, DenseNet-169 is one of the DenseNet architectures [14] that are available through Keras library - a Python library that allows development of DL models sequentially [1]. Note that the computations are performed on a NVIDIA Tesla V100 GPU that has 32 GB of RAM using TensorFlow among the available backend implementations of Keras [13].

Without prior training, the network performed well, however it necessitated 1000 epochs to be adequately trained. In order to reduce this number and thus the computation time to complete the training, we decided to investigate the benefit of transfer learning. For this reason the ImageNet dataset was included to take advantage of the pre-trained weights to reduce the number of epochs required, and consequently the computation time needed to complete the training. Transfer learning, which consists to apply a pre-trained model to a new task by providing a starting point for the training, is widely used in deep learning.

As ImageNet is a multi-class classification problem with 1000 classes, we had to adapt the classification layer of the adopted DenseNet network to reduce this number to 101 classes. Therefore, the output of the network's convolutional part is flattened and the 1000D fully-connected output layer replaced by a 101D one. To prevent overfitting dropout is added after flattening using a dropout rate of 50%. Like in the original DenseNet network, the activation function adopted in the classification layer is the softmax function.

At an early stage of this work an auto-encoder was used in order to be able to support various input image sizes. Indeed, we observed that the input images could have various sizes. Thus the idea was to add an auto-encoder, before the DenseNet network, to resize images in a uniform size while keeping as much information as possible. In this way, the auto-encoder was supposed to find the best match between an original input image and its version that would be fed into the classifier. However, when we used the images provided by the auto-encoder the classifier could only reach an accuracy level of 97%, and this preprocessing was also time consuming. As a result, we abandoned the auto-encoder and simply replaced it with a crop to get uniform image sizes of 224 * 224 pixels, which is the size required by the selected DensetNet architecture.

[1] https://keras.io/api/applications/densenet/

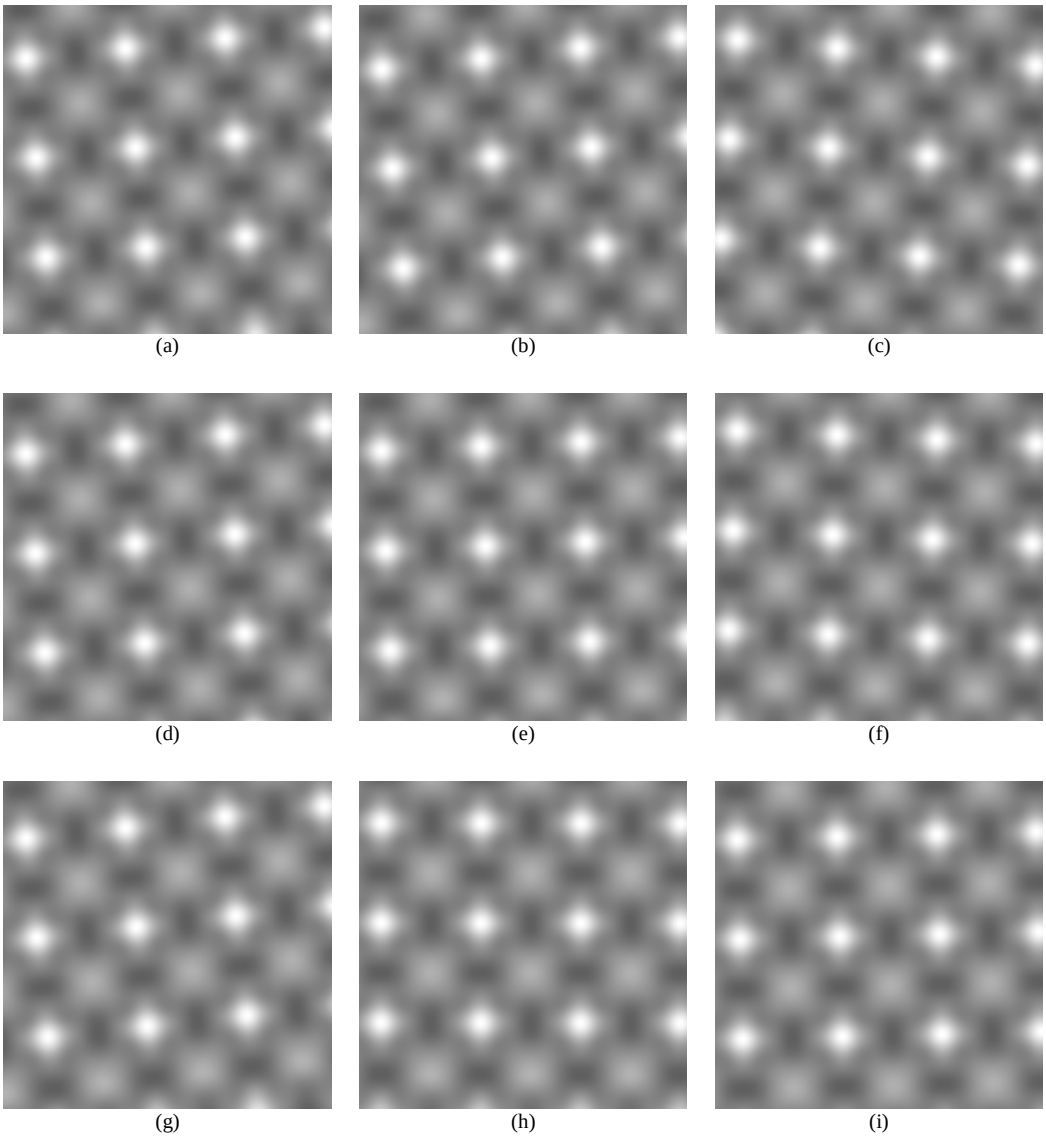

Figure 2. (a)-(c) Images of sight at distance 1.44 cm, (d)-(f) images of sight at distance 1.45 cm, (g)-(i) images of sight at distance 1.46 cm.

During the experiments, we observed that the DenseNet network was clearly overtraining after 500 epochs, which is why the "early stopping" technique was considered to prevent this. It consists in stopping the training once the model performance is no longer improving. To study this problem more precisely, 20 independent training runs were analyzed to check when overtraining occurs. On average, this was found to happen after 150 training epochs.

The adopted loss function is categorical cross-entropy, a loss function suited for problems with single label categorization. It applies a comparison between the per-class predictions' distribution in the output layer with the true distribution [17]. Furthermore, the Stochastic Gradient Descent (SGD) optimization was adopted with a learning rate of 1e-4 and a mini-batch size of 32. SGD is a well-known technique in DL that is used for weights optimization by minimizing the loss function [18].

Now we will give more details on the dataset and more particularly how the images were obtained. As previously explained in Section III, an holographic image generator was used to generate 101 classes of images. Moreover, image augmentation, which aims to increase the data volume by creating transformed versions of the original images [19], was applied during the training. It allowed us to obtain a larger data volume using the nearest-neighbor fill resizing technique while preprocessing and flipping horizontally and vertically the training images. Each class represents a different distance value from the target object. The distance values start with 0 with an increase of 100 micrometer for each class. Practically, it means that each class consists of 100 holographic images and thus that the dataset is

composed of 10,100 images. The dataset was split as per the 70-30 rule; 70% of the images were used for training, and the remaining 30% for the testing.

Multiple tests were conducted on the network to measure the accuracy and loss levels of the results while adapting what is needed to improve as much as possible the results. After fine-tuning, the network achieved an accuracy level of 100% when trained properly, at micrometer scale for the 101 classes, without needing to re-use the training images in the testing. Thus, among the different flavors of DenseNet investigated in [14], which have number of layers going from 121 up to 264 for ImageNet, the DenseNet with 169 layers is the shallowest one able to successfully classify any image. It means that the network with 121 layers is not able to provide enough discriminative feature maps before the classification layer.

V. CONCLUSION

In this paper, we have proposed a solution to determine the distance of a sight that is captured by a holographic camera. It benefits from deep learning techniques and consists in a CNN that is built on top of DensetNet. Finding this distance is indeed modeled as a classification problem where an input image must be classified within one class among 101. Training has benefited from transfer learning to reduce the number of epochs needed to complete it. The experimental results showed that the proposed network can predict at a micrometer scale with an accuracy level of 100% if trained properly, thus proving that DL techniques are more and more promising in the DH context. There are several possible avenues for future work that we would like to explore. First, further increasing the number of classes can be investigated. Second, the problem could be solved using regression instead of classification, which shall result in having continuous values of depth instead of discrete ones. Third, the impact on the results of changing the environmental setup such as luminosity may be evaluated, as well as predicting the distance of more than one recorded object having different distances.

ACKNOWLEDGMENT

This work was partially supported by the EIPHI Graduate School (contract ANR-17-EURE-0002). Computations have been performed on the supercomputer facilities of the "Mésocentre de Franche-Comté".

REFERENCES

[1] J. Sultana, M.U. Rani, and M.A.H. Farquad, "An extensive survey on some deep learning applications," Proc. CCODE 2019, Emerging Research in Data Engineering Systems and Computer Communications, Advances in Intelligent Systems and Computing book series (AISC, volume 1054), 2019, pp. 511-519, doi:10.1007/978-981-15-0135-7_47

[2] T. Tahara, X. Quan, R. Otani, Y. Takaki, O. Matoba, "Digital holography and its multidimensional imaging applications: a review, Microscopy," Volume 67, Issue 2, April 2018, pp. 55–67, doi.org/10.1093/jmicro/dfy007

[3] M.K. Kim, "Principles and techniques of digital holographic microscopy," SPIE Reviews, vol. 1, no. 1, Apr. 2010, doi:10.1117/6.0000006.

[4] I. Acharya and D. Upadhyay, "Comparative study of digital holography reconstruction methods," Procedia Computer Science, vol. 58, Dec. 2015, pp. 649-658, doi:10.1016/j.procs.2015.08.084.

[5] A. Hong, B. Zeydan, S. Charreyron, O. Ergeneman, S. Pané, M.F. Toy, A.J. Peruska, and B.J. Nelson, "Real-time holographic tracking and control of microrobots," IEEE Robotics and Automation Letters, vol. 2, no. 1, Jan. 2017, pp. 143-148,doi:10.1109/LRA.2016.2579739.

[6] T. Shimobaba, T. Kakue, and T. Ito, "Convolutional neural network-based regression for depth prediction in digital holography," Proc. International Symposium on Industrial Electronics (ISIE 2018), IEEE Press, June 2018, pp. 1323-1326,doi:10.1109/ISIE.2018.8433651.

[7] Z. Ren, N. Chen, and E.Y. Lam, "Automatic focusing for multisectional objects in digital holography using the structure tensor," Optics Letters, vol. 42, no. 9, 2017, pp. 1720-1723, doi:10.1364/OL.42.001720.

[8] H.A. İlhan, M. Dogar, and M. Özcan, "Digital holographic microscopy and focusing methods based on image sharpness," Journal of Microscopy, vol. 255, no. 3, 2014, pp. 138-149, doi:10.1111/jmi.12144.

[9] Z. Ren, Z. Xu, and E.Y. Lam, "Learning-based nonparametric autofocusing for digital holography," Optica, vol. 5, no. 4, 2018, pp. 337-344, doi:10.1364/OPTICA.5.000337.

[10] T. Pitkäaho, A. Manninen, and T.J. Naughton, "Performance of autofocus capability of deep convolutional neural networks in digital holography microscopy," Proc. Digital Holography and Three-Dimensional Imaging, 2017, OSA Technical Digest (online) (Optical Society of America, 2017), doi:10.1364/DH.2017.W2A.5.

[11] T. Pitkääho, A. Manninen, and T.J. Naughton, "Focus classification in digital holography microscopy using deep convolutional neural networks," Proc. Advances in Microscopic Imaging, International Society for Optics and Photonics 2017, OSA Technical Digest (online) (Optical Society of America, 2017), vol. 10114, July 2017, pp. 89-91, July 2017, doi:10.1117/12.2286161.

[12] Z. Ren, Z. Xu, and E.Y. Lam, "Autofocusing in digital holography using deep learning," Proc. Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXV, vol. 10499, 2018, doi:10.1117/12.2289282.

[13] S. Gu, M. Pednekar, and R. Slater, "Improve image classification using data augmentation and neural networks," SMU Data Science Review, vol. 2, no. 2, 2019, available at https://scholar.smu.edu/datasciencereview/vol2/iss2/1

[14] G. Huang, Z. Liu, L. Van Der Maaten, and K.Q. Weinberger, "Densely connected convolutional networks," Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), IEEE Publisher, July 2017, pp. 2261-2269, doi:10.1109/CVPR.2017.243.

[15] R. Kumar, "Adding binary search connections to improve densenet performance," Proc. 5th International Conference on Next Generation Computing Technologies (NGCT-2019), Feb. 2020, available at https://ssrn.com/abstract=3545071 or doi:10.2139/ssrn.3545071.

[16] M. Asmad Vergara, M. Jacquot, G. Laurent, P. Sandoz, "In-plane position and orientation measurement of a mobile target by digital holography," Proc. Latin America Optics and Photonics Conference (LAOP 2016), vol. LTh3C.3, August 2016, doi:10.1364/LAOP.2016.LTh3C.C

[17] M. Masrour, C. Lacaule, and A. Mohammad-Djafari, "Deep learning and artificial intelligence for the determination of the cervical vertebra maturation from lateral radiography," Entropy, vol. 21, no. 12, Dec. 2019, doi:10.3390/e21121222.

[18] E. Yazan and M.F. Talu, "Comparison of the stochastic gradient descent based optimization techniques," Proc. International Artificial Intelligence and Data Processing Symposium (IDAP 2017), IEEE Press, Sept. 2017, pp. 1-5, doi:10.1109/IDAP.2017.8090299.

[19] S. O'Gara and K. McGuinness, "Comparing data augmentation strategies for deep image classification," Proc. Irish Machine Vision and Image Processing (IMVIP 2019), 2019, available at https://api.semanticscholar.org/CorpusID:204845142.