Implementation of Convolutional Neural Networks in a Virtual Optical System for Wavefront Detection with Potential Application in Visual Optics

Andrés Osorno-Quiroz,*[1] Walter Torres-Sepúlveda,[2]

[1]*Programa de Ingeniería de software y datos, Facultad de Ingeniería y Ciencias Agropecuarias, Institución Universitaria Digital de Antioquia, Medellín, Colombia*

[2]*Programa de Ciencias Ambientales, Grupo de Investigación en Innovación Digital y Desarrollo Social (INDDES), Facultad de Ciencias y Humanidades, Institución Universitaria Digital de Antioquia, Medellín, Colombia*

Received June 19, 2024; accepted September 24, 2024; published September 30, 2024

Abstract— In this work, the effectiveness of the convolutional neural network architectures AlexNet and a new proposed ResNet-based architecture are compared for the detection of optical aberrations in typical images from a Hartmann-Shack sensor, within the range of values associated with the average aberrations of a real eye. Both neural networks are trained with a dataset built from a virtual optical system containing more than 44,000 training, validation, and testing images. The results demonstrated that both neural networks were able to accurately predict the typical aberrations simulated for a real eye, with better performance for the proposed ResNet CNN. Compared to traditional methods such as the centroid detection method in Hartmann-Shack images, this artificial intelligence-based approach presents itself as an effective and promising alternative for aberration detection where there are no such restrictive conditions as dynamic range, making this methodology and the proposed ResNet potentially applicable in fields such as adaptive optics and ophthalmology.

Controlling the optical quality of physical and optical systems is crucial in technological, scientific, and current medical developments. Essentially, accurate measurement of optical quality is linked to adequately diagnosing an optical system's problems and clearly to its subsequent correction. This line of thinking is evident, for example, in fields such as optometry and ophthalmology, where the best correction for a visual condition is tied to the quality of the diagnosis made.

Currently, detecting the optical quality of an optical system is made by studying its optical aberrations, which can be determined from images obtained by a Hartmann-Shack sensor (HSS) [1]. These aberrations, including second-order (defocus and astigmatism) up to high-order ones such as spherical aberration, coma, and trefoil, are the causes of image quality degradation in applications such as telescope construction and measurement of optical and visual quality in real eyes.

Traditionally, methods for detecting aberrations have relied on image segmentation and centroid positioning techniques [2], which is often a laborious and error-prone process. In this sense, the rise of deep learning has enabled the initial proposal of some models aimed at improving the measurement of optical aberrations in different fields. For example, Guo *et al.* in [3] applied backpropagation neural networks to estimate Zernike coefficients using centroid displacements on an HSS, comparing the results with traditional methods. Barwick proposed a hybrid astigmatic wavefront approach with post-processing using neural networks [4], while Li *et al.* employed neural networks to calculate centroids in HSS images under extreme noise conditions [5]. Hu et al. applied a modified U-net to directly reconstruct the wavefront distribution from HSS patterns with a specific root-mean-square (RMS) wavefront error [6].

All of the previous methods demonstrate that the neural networks can improve the performance of the HSS. However, particularly in the area of visual optics, it is common for traditional approaches based on centroid detection and measurement of spot displacements in the HSS image to be still applied, which subsequently allow for the calculation of Zernike coefficients, used as a standard for wavefront reconstruction and therefore for determining the optical quality of a system.

Taking advantage of the promising advances in the field of deep learning over the last decade, this article proposes the implementation of two CNNs: AlexNet and a customized ResNet architecture. The goal is to compare and evaluate their performance on computationally simulated (virtual optical system) typical images from an HSS obtained over a range of optical aberration values comparable to those found on average in human eyes [7].

The AlexNet CNN architecture, consisting of 5 convolutional layers and 3 fully connected layers, is implemented in this work to directly predict the 18 most relevant Zernike coefficients (3 to 20 in OSA notation) for characterizing wavefront aberrations in the human eye from HSS images. The first convolutional layer filters the input with 32 kernels of size 5×5, followed by a second layer with 32 3×3 kernels. The following 3 convolutional layers use 64 3×3 kernels each. The flattened results are then passed through fully connected layers of 512, 170, and finally, 18 neurons that output the Zernike coefficient regression values. This architecture is similar to Li et al.'s work [5] but adapted to predict only the human eye's representative aberration modes, enabling an end-to-end data-driven approach avoiding explicit centroid detection and feature extraction used in traditional methods.

*E-mail: andres.osorno@est.iudigital.edu.co

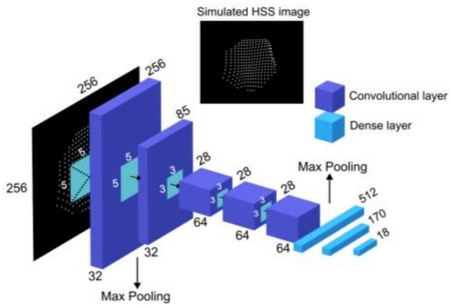

Fig. 1. Architecture of the implemented CNN AlexNet.

On the other hand, the ResNet CNN architecture, as shown in Fig. 2, begins with an initial block where we find two convolutional layers with 64 filters, each one with a kernel size of 5×5. In this same block, we find max-pooling layers, batch normalization, and the activation function, which is again ReLU (Rectified Linear Unit). Subsequently, we find the residual blocks 1, 2, and 3, in which two convolutional layers with 64, 128, and 256 filters are applied in this order with identity mapping. The output of the first convolutional layer is passed through Batch Normalization and ReLU. The skip connections, or residual connections, are applied within these blocks, allowing the original input information to "skip" directly through the convolutional layers and combine with the output of those layers. This facilitates the gradient flow and optimizes the training process in deep networks. At the end of the network, we have a dense layer with 18 neurons that provides the predicted 18 Zernike coefficients. To our knowledge, this exact architecture has not yet been implemented to analyze HS images for any optical system, specifically for images with coded aberrations associated with the human eye.

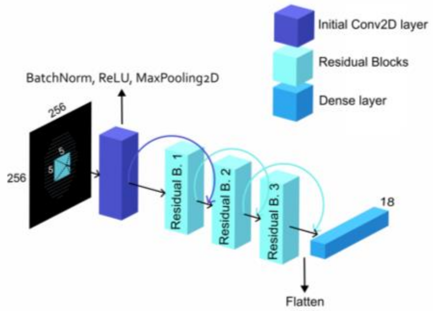

Fig. 2. Architecture of the ResNet-based implemented CNN.

The evaluation of both models was performed using the mean square error (MSE) cost function, taken as a standard for the evaluation of optical systems through the optical aberrations represented by the Zernike coefficients. ADAM (Adaptive Moment Estimation) was also implemented as the optimization function for both networks.

For the creation of the dataset that would serve as input for training the networks, a custom-made virtual optical system was implemented, based on Fourier optics theory and light propagation models, which allowed recreating the spot pattern map with the respective displacements associated with a list of Zernike coefficients. That is, it allowed obtaining the displacements induced by the local slope of the wavefront, represented by the expansion in Zernike coefficients, for a total of 21 Zernike terms (Z_j), eliminating from there the piston, tip, and tilt (j=0, 1 and 2, in a single index notation) given a list of Zernike coefficients with 18 terms.

The images consisted of two groups: one with pure aberration terms, where only one of the 18 coefficients was non-zero. In contrast, the other group consisted of images with random combinations of the 18 Zernike coefficients. The ranges used were [-7,7] μm for oblique astigmatism, defocus, and horizontal astigmatism (j=3, 4 and 5), while for the higher-order aberrations, coefficients with j=6, ..., 20 in single index notation, the range was [-1,1] μm. These ranges were selected to evaluate the variability of the different Zernike coefficients within the average human eye.

In total, 44884 images were generated for training, along with 8192 validation data sets and 2048 test data sets for both models. The test and validation images are generally different from each other. The original dimensions of the images were 1280×1024, but they were resized to 256×256 to achieve a shorter training time since this resizing does not compromise the model accuracy. The simulated pixel size was assumed to be 5.2 μm, the pupil size was 6 mm, and the wavelength was 0.532 μm.

This dataset configuration provides a comprehensive basis for training and evaluating the performance of the convolutional neural network models on Hartmann-Shack wavefront sensing tasks. The extensive training set of 44884 images spanning a diverse range of aberration patterns enables the models to effectively learn the complex mapping between spot displacement patterns and Zernike coefficients.

Figure 3 shows the training curves. The MSE is displayed as the loss function on the vertical axis, and on the horizontal axis, the corresponding training epochs are shown, which were limited to 50 epochs. From the figure, it is clear that there is a decreasing behavior of the cost function for both CNNs. However, it can be seen how the proposed ResNet in this work always shows above the training line, in contrast to AlexNet, thanks to the skip connections that preserve the gradient norm on the backward path. This facilitates the preservation of the gradient norm, keeping this network stable and avoiding overfitting.

The total training time using AlexNet with the aforementioned dataset was 42 minutes to complete 50 epochs with a batch size of 32. This network was evaluated with the test data, and an error percentage close to 3.86% was obtained, and the Root Mean Squared Error (RMSE) was 0.073 μm. The RMSE was implemented to compare the actual Zernike coefficients with the predicted ones in order to give an objective value in the predictive capabilities of the CNNs.

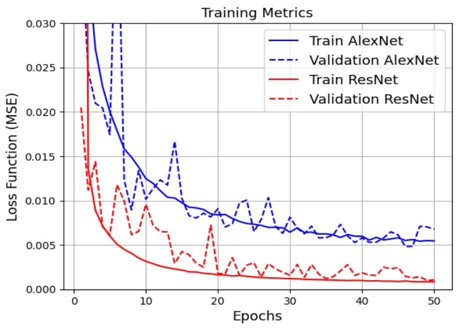

Fig. 3. The training curve obtained for both CNNs AlexNet and ResNet.

On the other hand, the training time using ResNet was 2.5 hours. The percentage of error decreases considerably in relation to AlexNet, positioning itself at 3.02% with an RMSE of 0.031 µm, demonstrating better performance in terms of prediction capability, although with a longer training time, which is not a crucial factor for this work.

This result is evident in Figs 4(a) and (b), where the Box-Plot is presented using the same test data set. As is evident, the dispersion of the differences is closer to zero for ResNet than for AlexNet (note the same scale). This conclusion is supported by the phase maps of a single combination of Zernike coefficients. As can be seen, compared to the original phase profile, ResNet presents a more remarkable similarity; however, from a graphical point of view, it can be said that both networks deliver an excellent representation of the wavefront.

Figure 4 validates that both CNNs are suitable for expanding the capabilities of an HSS for measuring optical aberrations. In particular, it can be noted that it has not been necessary to impose restrictions regarding the maximum aberrations that can be processed by CNNs, contrary to the case of the traditional implementation of an HSS, where the dynamic range is a crucial factor and limits its use in contexts where aberrations are high, as can happen in eyes with high myopia or aberrations over telescope images, affected by significant changes in more active atmospheres.

Beyond performance metrics, ResNet's adaptability to different types of images adds a valuable nuance to these conclusions. Thus, ResNet's ability to handle random images more effectively expands its applicability, highlighting its versatility in optical applications. These results agree with those of Zhang *et al.* [8], where a ResNet-based CNN (but different from our proposal) was also implemented. While other similar approaches exist, they differ significantly from our proposal. Jian et al. [9] use a more complex ResNet-34-based CNN, focusing only on astigmatism and coma aberrations for a single lens using diffraction patterns, excluding other aberration types. In contrast, Zhan *et al.* [10] also employs a complex ResNet-34 based CNN but with a zonal approximation, unlike our modal methodology implemented with the HSS. Their approach obtains aberration compensation directly from the point spread function without determining wave aberrations, which is the primary goal of our work.

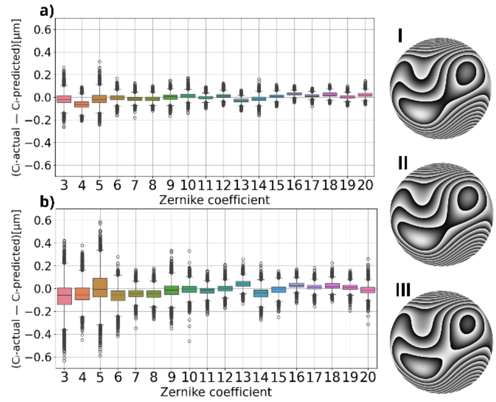

Fig. 4. Comparison of ResNet and AlexNet predictions. (a) and (b) Box plot with the difference between the actual Zernike and the Zernike predicted by the CNN (µm). (I) Actual phase map of a random image. (II) ResNet prediction of the actual phase map. (III) AlexNet prediction of the actual phase map.

In summary, it can be concluded that the results support not only the applicability of CNNs for problems associated with the detection of typical eye aberrations but also show the superiority of the proposed custom-made ResNet architecture, which has not yet been implemented for problems associated with visual optics. By virtue of this, when dealing with HSS images, ResNet is the best option to implement due to its residual blocks with skip connections. From a perspective, it is essential to bring these strategies closer to experimental image sets, where other challenges arise. It may modify the final performance of CNN models, including images with background speckle noise or dynamic changes in pupil sizes.

This research was supported by grants from the Joven Investigador program at Universidad EAFIT and by Universidad de Antioquia (Convocatoria CODI 2021-2022). Walter Torres-Sepúlveda also acknowledges support from the Fondo Francisco José de Caldas (Colciencias-Colombia, grant 727 of 2015).

References

[1] J.C. Ochoa-Tabares, E. Hernández-Quíntela, N.D.C. Ruiz-Quintero, R. Naranjo-Tackman, Revista Mexicana de Oftalmología **86**, 25 (2012).
[2] J. Li, Y. Gong, X. Hu, C. Li, Chin. J. Lasers **41** (2014).
[3] H. Guo et al., Opt. Expr. **14**, 6456 (2006).
[4] S. Barwick, Opt. Lett. **34**, 1690 (2009).
[5] Z. Li, X. Li, Opt. Express **26**, 31675 (2018).
[6] T.B. DuBose, D.F. Gardner, A.T. Watnik. Opt. Lett. **45**, 1699 (2020).
[7] S. Plainis, I.G. Pallikaris, J. Mod. Opt. **55**, 759 (2008).
[8] H. Zhang, J. Zhao, H. Chen et al., *Sensors*. **24**, 2728 (2024).
[9] B. Zhang, J. Zhu, K. Si, W. Gong, Front. Phys. **8**, 621966 (2021).
[10] J. Jiang, X. Liu, Y. Chen, S. Gao, Y. Liu, Y. Jiang, AIP Adv. **13**, 105322 (2023).

