

Article

Assignment of Focus Position with Convolutional Neural Networks in Adaptive Lens Based Axial Scanning for Confocal Microscopy

Katharina Schmidt ^{1,2} , Nektarios Koukourakis ^{1,2,*}  and Jürgen W. Czarske ^{1,2,3,4} 

- ¹ Laboratory of Measurement and Sensor System Technique, TU Dresden, Helmholtzstrasse 18, 01069 Dresden, Germany; katharina.schmidt1@tu-dresden.de (K.S.); juergen.czarske@tu-dresden.de (J.W.C.)
² Biomedical Computational Laser Systems (BIOLAS), TU Dresden, 01069 Dresden, Germany
³ Cluster of Excellence Physics of Life, TU Dresden, 01069 Dresden, Germany
⁴ Faculty of Physics, TU Dresden, 01069 Dresden, Germany
* Correspondence: nektarios.koukourakis@tu-dresden.de

Abstract: Adaptive lenses offer axial scanning without mechanical translation and thus are promising to replace mechanical-movement-based axial scanning in microscopy. The scan is accomplished by sweeping the applied voltage. However, the relation between the applied voltage and the resulting axial focus position is not unambiguous. Adaptive lenses suffer from hysteresis effects, and their behaviour depends on environmental conditions. This is especially a hurdle when complex adaptive lenses are used that offer additional functionalities and are controlled with more degrees of freedom. In such case, a common approach is to iterate the voltage and monitor the adaptive lens. Here, we introduce an alternative approach which provides a single shot estimation of the current axial focus position by a convolutional neural network. We use the experimental data of our custom confocal microscope for training and validation. This leads to fast scanning without photo bleaching of the sample and opens the door to automatized and aberration-free smart microscopy. Applications in different types of laser-scanning microscopes are possible. However, maybe the training procedure of the neural network must be adapted for some use cases.

Keywords: adaptive optics; axial scanning; confocal microscopy; deep learning; neural networks; transfer learning



Citation: Schmidt, K.; Koukourakis, N.; Czarske, J.W. Assignment of Focus Position with Convolutional Neural Networks in Adaptive Lens Based Axial Scanning for Confocal Microscopy. *Appl. Sci.* **2022**, *12*, 661. <https://doi.org/10.3390/app12020661>

Academic Editors: Hao-Chung Kuo and Andrés Márquez

Received: 10 November 2021

Accepted: 6 January 2022

Published: 10 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Confocal microscopy (CM) is frequently used in life sciences to investigate biological samples, such as tissues and cells, where high spatial resolution is necessary to obtain information about small structures of the sample. The main advantage of CM is the optical sectioning capability, introduced by a confocally aligned pinhole [1,2]. However, CM is a point-wise measurement technique and thus scans in three dimensions are required to acquire three-dimensional information. To accomplish an axial scan, several methods are known. In stage scanning, the sample is moved through the constant focus by a stage as introduced in [3], though motion artefacts in the scan may occur. Another commercially available possibility for axial scanning is a piezo-driven objective holder [4]. Alternatively, deformable mirrors (DM) may also be used for axial scanning [5] and aberration correction [6,7]. In recent years, another general approach based on adaptive lenses (AL) was introduced [2,8–10]. By tuning the voltage applied to the AL, the focus shifts axially without the need for any mechanical movement. DMs are easy to control and very powerful, however, they usually have a smaller upstroke compared with AL and require a folded beam, which leads to a bulky optical setup in general. ALs do not need a folded beam because they are translucent elements, so the setup is more compact. There are various realisations of ALs. The most popular concepts include tunable acoustic

gradient (TAG) lenses [11], liquid crystal lenses [12], lenses actuated by electrowetting [13], and deformable lenses with piezo actuators [10,14,15].

For axial scanning, the focal spot has to be shifted in the axial direction through the sample. To achieve such a shift, the AL induces a difference of the optical path length (*OPL*) when a beam is passing through the lens. Here, the *OPL* is defined according to Equation (1), where n is the refractive index of the lens material and s is the length of the path through the lens:

$$OPL = n \cdot s \quad (1)$$

TAG lenses achieve this by using sound energy to change the refractive index of the medium inside the lens [11], while the length s is constant. The alternative to a change of the refractive index is to adjust the local thickness of the lens by the deformation of the AL membrane with piezo actuators [14]. Such lenses enable high-speed axial scanning without the need for mechanical movement of the stage or microscope objective. However, environmental influences have a huge impact on the behavior of the AL. Therefore, it is necessary to monitor the behavior of the lens while scanning and to always be aware of the current position of the focal spot.

The control of AL with higher degrees of freedom, such as the lens with spherical aberration correction capabilities introduced in [14], becomes more challenging as the axial position of the focal plane varies based on unknown or non-linear dependencies. If such an AL is inserted into an optical setup closed-loop control supported by wave, front measurements can be used for control. There, the voltage is iteratively refined until the desired behaviour is reached. However, this is time-consuming, and the sample is exposed to the laser for a significant amount of time, hence, photobleaching might pose a problem.

To overcome these limitations we introduce a computational approach to estimate the current position of the focal plane while scanning and without iterations in the scanning process. The strategy is to monitor the behaviour of the scanning AL with a camera and assign the axial position to the recorded measurement point.

In several use cases, techniques of deep learning were able to solve non-linear problems better than a classical closed loop control [16]. Neural networks as a method of deep learning are able to consider a huge amounts of information with unknown importance [17]. Furthermore, a neural network avoids time-consuming iterations once it is trained, and photobleaching is expected to be strongly reduced.

For focal length estimation, different approaches based on images are implemented in the form of convolutional neural networks (CNN). It is possible to deal with the above-mentioned issue by calibration using specific patterns on a single image [18–20] or the usage of an RANSAC-based algorithm [21]. All these approaches have in common that a specific pattern must be visible on the image, so that the focal length is calculated based on the geometric features of this pattern. In addition, some techniques require more than one image to estimate the focal length [18], so no single shot estimation is possible. There are also newer approaches with deep learning in the form of neural networks to estimate the focal length from a single image without specific patterns [22–25]. However, all the above-mentioned methods were only applied on pictures of subjects such as landscapes, persons, animals, cells, or synthetically generated images. In the case of microscopy, it is beneficial to be independent from the type of sample. To our knowledge, there is currently no neural network to estimate the focal shift introduced by adaptive optics based on images without a subject on it available [22–25].

2. Experimental Setup

For training the CNN, we use experimental data. Therefore, we use a home-built confocal microscope setup, as shown in Figure 1. The laser source is a Thorlabs CPS532 laser with 4.5 mW power operating at a wavelength of $\lambda = 532$ nm. The beam passing the AL and the lenses L2 and L3 is divided by a beamsplitter into the measurement and the monitoring beams. The monitoring beam is directed onto a camera through lens L5, where the input image for the CNN is captured. L4 (Ashperical Lens 354453, Thorlabs Inc.,

Bergkirchen, Germany) focuses the measurement beam into the region of interest. Here, an aspherical lens was used as a front lens.

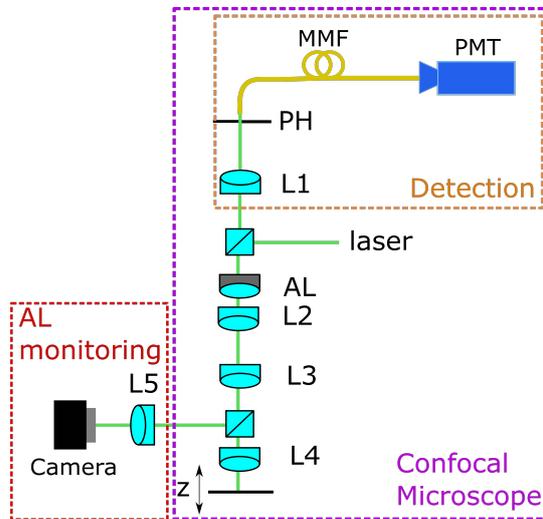


Figure 1. Confocal microscopic setup with adaptive lens, AL= adaptive lens, L = lens, MMF = multi mode fibre, PMT = detector, PH = pinhole.

For calibration, a mirror which is located on a movable stage (stage controller XPS-Q8 and motorized actuators LTA-HS, Newport Corporation, Irvine, CA) is used as a sample. This stage is movable in three axes, but for the data acquisition, it has only been moved in the axial z-direction. The lenses L2 and L3 form a 4f-system that images the AL close to L4, which minimizes aberrations of the adaptive lens but still offers a high tuning range as described in [26].

The reflected light goes back, passing the 4f-system, is focused through the pinhole ($d = 5 \mu\text{m}$ from Thorlabs Inc.), and is measured with a detector (HCA-S Femto, Berlin, Germany). The light originating from out-of-focus regions is blocked by the pinhole. When the mirror is driven through the focus, the intensity distribution measured on the detector gives both the focus position and the axial resolution. If the voltage applied on the AL is changed, the peak of the intensity is shifted, and the focus position can be determined. The working principle of the AL is shown in Figure 2.

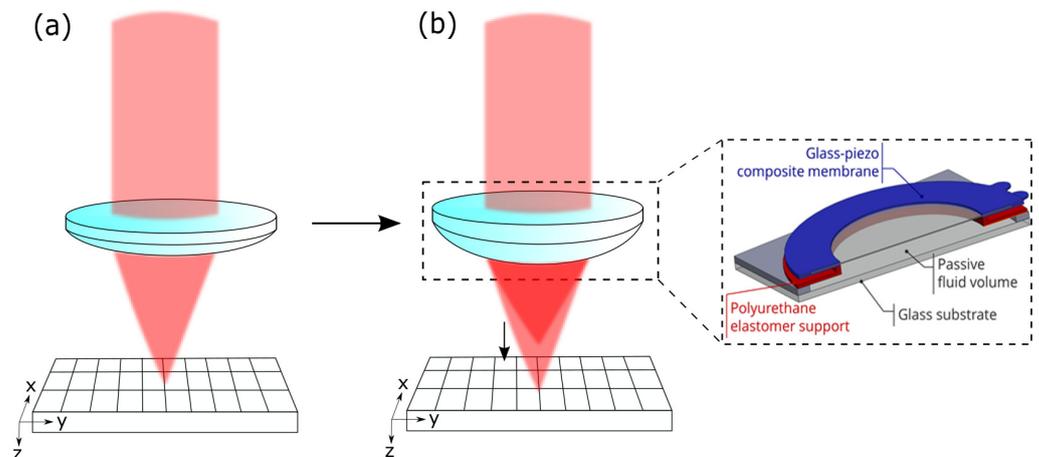


Figure 2. (a) Conventional lens with fixed focal spot. (b) AL enables axial scanning by shifting the focal spot; this is achieved by a deformation of the membrane of the lens.

Assuming the lateral resolution in a confocal microscope to be defined as in Equation (2) and the axial resolution as in Equation (3) from [27]:

$$\Delta x, \Delta y = \frac{0.51 \cdot \lambda}{NA_{AL}} \quad (2)$$

$$\Delta z = \frac{0.64 \cdot \lambda}{n - \sqrt{n^2 - NA_{AL}^2}}. \quad (3)$$

With $\lambda = 570$ nm as the average wavelength of emission and excitation (defined by Equation (4)), n is the refractive index and NA_{AL} is the numerical aperture. NA_{AL} can be derived from Equation (5), here, this is the numerical aperture, which changes when a voltage is applied to the AL (induced by a change of the focal length f_{obj}) [27]:

$$\lambda = \sqrt{2} \cdot \frac{\lambda_{ex} \cdot \lambda_{em}}{\sqrt{\lambda_{ex}^2 + \lambda_{em}^2}} \quad (4)$$

$$NA_{AL} = n \cdot \sin(\alpha) = n \cdot \sin(\arctan(\frac{f_{obj}}{r})) \quad (5)$$

Here, the refractive index is assumed as $n = 1$, r is the half size of the spot in front of L4, and f_{obj} is the focal length behind L4, influenced by the AL ($f_{obj} = 4.61$ mm while 0V applied). The spot size of the laser beam is $d = 10$ mm, so r is 5 mm; thereby, NA is calculated as 0.68. So, the lateral resolution becomes $0.4 \mu\text{m}$, and the axial resolution is $1.4 \mu\text{m}$. The radius of the Airy disk can be calculated according to Equation (6) [27]:

$$r_{Airy} = \frac{0.61 \cdot \lambda}{NA_{AL}}. \quad (6)$$

This leads to a radius of the Airy disk of $0.5 \mu\text{m}$:

$$r_{PH,proj} = \frac{r_{PH}}{M}. \quad (7)$$

$$AU = \frac{r_{PH,proj}}{r_{Airy}}. \quad (8)$$

M is the overall magnification of the setup, which can be stated as 10.7, r_{PH} is the size of the pinhole. Thereby, $r_{PH,proj}$ and the Airy Unit AU are according to Equations (7) and (8) from [27] $0.46 \mu\text{m}$ and $0.89 \mu\text{m}$.

The standard deviation of the focus position upon repeated application of a voltage on the adaptive lens at changing environmental conditions was found to be in the order of $50 \mu\text{m}$. To obtain a ground truth (gt) value for the current axial position of the focal spot, when a certain voltage is applied to the lens, the following procedure was performed. After applying a voltage to the AL, the stage was moved axially, and simultaneously, the intensity at the detector was recorded. The axial position of the focal spot was afterwards calculated as the position of the stage when the maximum intensity was measured at the detector. The axial resolution of the microscope is therefore crucial for the accuracy of the gt value.

3. Training of the Neural Network

Our network uses images captured from the monitoring beam of our CM setup. Hence, the position estimations are independent from the investigated sample. We use only experimental data to build a dataset for the network training, and this leads to an exceptionally small dataset.

Each data element in the dataset consists of an image from the illumination beam and the gt value of the axial focal spot position. To obtain a dataset for training, validation, and testing, for each element, the voltage AL is changed and the capturing procedure is repeated.

We use the AL introduced in [15] and apply voltages between 0 V and 20 V, resulting in an axial scanning range of 230 μm.

The CNN from Workman et al. [22] showed good performance in focal length estimation by using the feature extractor from AlexNet [28]. We also adapted this extractor into our model with some changes in the architecture. This results in a modified AlexNet with a different concept than the original network. To decrease the number of weights and speed up the training, we dropped some layers (including the dropout layers) from the original extractor. At the end of our architecture, we inserted one linear layer with a single output to obtain the focal length as a single number. An overview of the resulting architecture is shown in Figure 3.

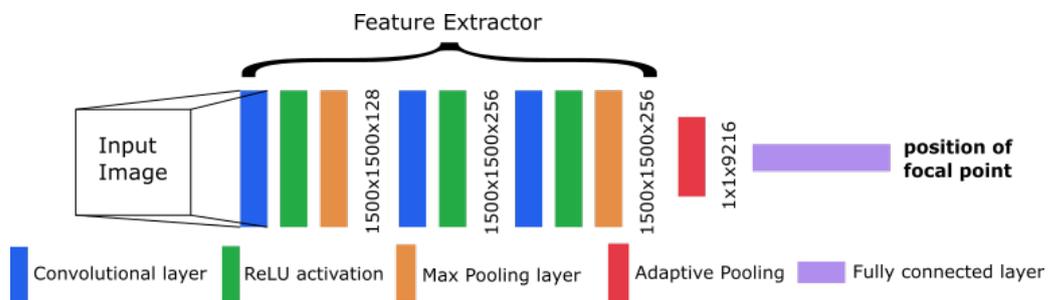


Figure 3. Overview of the convolutional neural network (CNN) architecture.

3.1. Implementation Details

Our implementation is written with PyTorch Lightning framework [29]. This framework improved the comprehensibility of our code and simplified the training process. As a loss function, we decided to implement the mean squared error (MSE) between the estimated and ground truth focal length. The expression for MSE is given in Equation (9):

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_{estimated} - x_{gt})^2 \tag{9}$$

where $x_{estimated}$ is the focal length estimated by the network and x_{gt} is the annotated gt focal length, while N is the number of elements used for the calculation. This loss calculation has the advantage that large deviations of the estimated value from the ground truth value have a larger weight. Small deviations are weighted as less intense but still allow the loss to converge over epochs.

To obtain a better understanding for the difference between the gt and the estimated value as an alternative, the mean absolute error (MAE) can be used. The MAE is defined as average difference, see Equation (10):

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_{estimated} - x_{gt}| \tag{10}$$

To clarify the test loss, we use the MAE in addition to the MSE as a metric. The results on the test-dataset are provided in Table 1.

3.2. Training Procedure

For training, we used an Adam optimizer and an initial learning rate of 1.995×10^{-4} . This learning rate was calculated by the learning rate finder from PyTorch Lightning framework [29]. The optimal learning rate is at the point of highest initial descent of training loss and provides an optimal start of the training process. The loss function over

the 363 epochs of the first training is shown in Figure 4. The curve is rapidly decreasing in the first epochs and afterwards is decreasing more slowly. This behavior is expected and indicates a successful training process. The split of the dataset into the training and validation datasets is 60:40, with 37 elements in total. Each element contains one image and one ground truth value of the axial focal spot position.

Table 1. Results of the trained CNN on the test dataset; TF = transfer learning; *MSE* = mean square error; *MAE* = mean average error; *STD* = standard deviation.

	<i>MSE</i> [μm^2]	<i>MAE</i> [μm]	<i>STD</i> [μm]	Accuracy [%]
1st Dataset	53.61	6.5	3.88	75
2nd Dataset before TF	48,486.00	213.36	62.85	0
2nd Dataset after TF	69.59	8.15	2.01	75

Additionally, we used four elements as test-set; please take note that there was no data augmentation added to the test set, and the elements were completely unknown to the CNN. The dataset is comparably small because all elements in the dataset were annotated by hand. Often, synthetic data are created by simulations to obtain large datasets for training. However, here, we had no model available that could adequately produce synthetic data and take into account experimental and environmental influences. However, as it is shown in Section 4, even on a small dataset a successful training is possible and can achieve satisfying results.

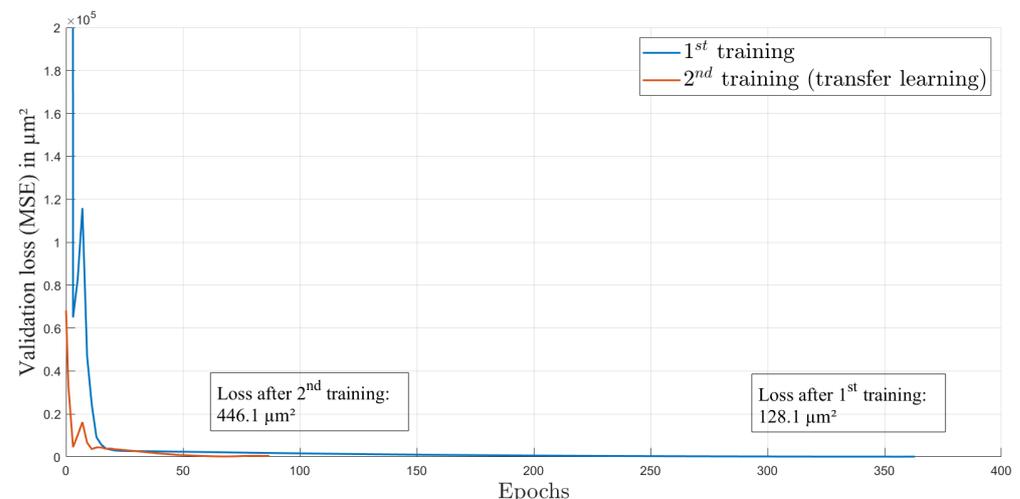


Figure 4. Loss over epochs for both training processes.

However, increasing the used data may offer slightly better results, but in future steps, it is planned to face highly increased difficulties, as more complex adaptive lenses will be used, with two or more input voltages, e.g., for additional aberration correction. Then, phase measurement has to be included into the setup, which will strongly increase the complexity. Thus, the dimension of the dataset will also be increased to address the extended set of features. These features will also include a high variety of aberrations. Hence, it is fundamental to keep the complexity at this stage as low as possible, thus the used number of data was suitable to reach the targeted estimation quality. An adequate CNN-based solution is the only way to make the adaptive optical component-based microscopy competitive.

4. Experiment and Results

In the last epoch of training (epoch 363), our model estimates the focal length on the validation data as *MSE* of $128 \mu\text{m}^2$ from the corresponding ground truth on the validation

data. Furthermore, random noise was added as data augmentation to the validation and training datasets, but for testing, a small dataset without data augmentation was used. For data augmentation, Gaussian noise from [30] was used. Here, noise in the form of counts was added to each pixel per image. The noise was sampled from a normal distribution with a mean of zero counts and a standard deviation of eight counts for the first training process. Thereby, it was possible to enhance the dimension of the training and validation datasets to 13,431 images. On this test dataset, we end up with a test loss of $6.5 \mu\text{m}$, calculated as *MAE*. This test loss depends on the overall performance of the network and its hyper-parameter but also on the quality and quantity of the training data. The training was performed on a system consisting of AMD Ryzen 9 3900X 12×3.80 GHz (CPU) and Corsair DIMM 128 GB DDR4-2666 Quad-Kit (RAM) and GeForce RTX 2080 Ti Blower (GPU). Here, the first training took 69 min.

To prevent our model from overfitting and to achieve a good generalization, we used the following strategies:

- Early stopping after four validation epochs without decreasing validation loss;
- Data augmentation (random noise) [30].

To assess the results of an experimental application of the CNN in combination with the AL, measurement uncertainties should be considered first. As potential sources of uncertainties in the setup, we identified the movement stage and random measurement uncertainties while capturing the gt values and camera parameters.

The motorized stage has an unidirectional repeatability of $0.25 \mu\text{m}$ and provides minimum incremental motion of $0.1 \mu\text{m}$ [31]. These tolerances are far below cell size ($10 \mu\text{m}$), thus, we neglect these influences. To minimize random measurement uncertainties with unknown origin, we used the average of 10 axial scans to obtain the gt value of the position of the focal plane. The camera parameters have a huge impact on image quality criteria such as the saturation of pixels, etc. The quality of the images for the CNN is an important factor for the learning process. Additional changes in the images through unsuitable camera parameters might make a correct estimation more difficult to achieve. To deal with this problem, we used the automatic mode for camera usage to avoid over saturation and other issues.

For complex specimens, sample-induced aberrations may be an issue that could degrade the performance quality, as the focus position estimations of the CNN are made independent of the sample.

To investigate the quality of the estimations made by the CNN, loss functions and accuracy calculations are a common method of evaluation. In Figure 5a, the gt values and estimated values for each data element in the training, validation, and test sets are shown. Here, the estimations of the trained CNN on all available data elements are shown. Of course, the elements for testing were not included in the training process; there, only training and validation elements were used. The elements from the test set are marked with a black star. Our main goal was to achieve a precision of $10 \mu\text{m}$ (average size of thyroid cells in zebrafish according to [14]), and to illustrate this, a shade around the gt with this width is added. From this image, one can see that this goal is mainly achieved for the middle elements, while the elements in the beginning and the end are differing stronger. Elements at the beginning and the end are linked to a very low or very high voltage and so to the clearest distortion in the input image for the CNN. It is possible that learning features in the middle of a range is easier for the CNN, and this might depend on the camera parameters used and the possible occurrence of over saturation at the end of the range. However, due to the test loss of $6.5 \mu\text{m}$ (*MAE*), our CNN obviously shows a good performance inside the desired range of tolerance. The statistical results on the test datasets are summarized in Table 1. Here, the accuracy of the results on the test dataset consisting of four elements is also reported. Each estimation, which differs less than $10 \mu\text{m}$ from the ground truth value, is considered as a correct estimation.

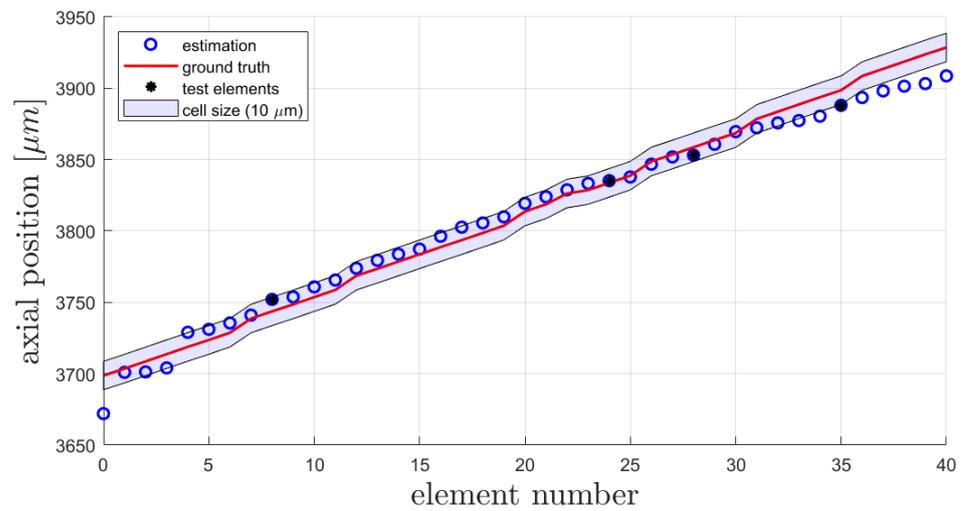
After several weeks and some lens adjustment, another dataset from the same setup and also consisting of 41 data elements was recorded to investigate the performance of the network on data after the AL was influenced by environmental circumstances such as changing ambient temperature and pressure. We used four random elements of this dataset for testing with the trained network and reached a test loss of $213.36 \mu\text{m}$ (*MAE*), see Table 1. As shown in Figure 5b (blue circles), the estimations do not match the corresponding gt values anymore. This is caused by the intentional misalignment of the setup, leading to different gt values compared to the first dataset. Here, the elements from the test dataset are marked in black. The test elements were not included in the training or validation process. We used transfer learning [32] to improve the performance, and again, data augmentation was used to extend the dataset. As in the first training process, Gaussian noise was added to the images, but here, the noise was sampled from a normal distribution with a standard deviation of 10 counts. This enables the extension of the training and validation data to 3219 images. In the method of transfer learning, a trained CNN is again trained with a slightly different dataset on top. This enables the CNN to recognise input elements with small discrepancy compared to the first dataset. After 87 epochs of transfer learning with heavier data augmentation than in the first training, the CNN achieved a test loss of about $8.15 \mu\text{m}$ (*MAE*). The whole process of transfer learning was conducted with the second dataset, which was recorded several weeks after the first dataset. The validation loss while training is provided in Figure 4. The improvement in correctness of the network estimations can also be seen in Figure 5b, where the estimations after transfer learning are displayed with green circles and the test elements are marked in black. It is obvious that some pattern of estimation distribution from before transfer learning are still visible in the estimations after transfer learning. By comparing the mean difference (*MAE*) of gt and estimation before transfer learning ($213 \mu\text{m}$) and after transfer learning ($8 \mu\text{m}$), a huge improvement is shown. Furthermore, the standard deviation (*STD*) of the difference also decreased from $62 \mu\text{m}$ before transfer learning to $2 \mu\text{m}$ afterwards. The transfer learning process took 16 min on the above-described system.

As an exemplary application, a scan through a fluorescent specimen was performed. In Life Sciences, the investigation of fluorescent specimen by using confocal microscopy is a principal task [33]. To prepare a fluorescent sample fluorescent beads were placed onto a mirror, followed by a microscope slide where on top again fluorescent particles were located. A scheme of this sample is shown in Figure 6a.

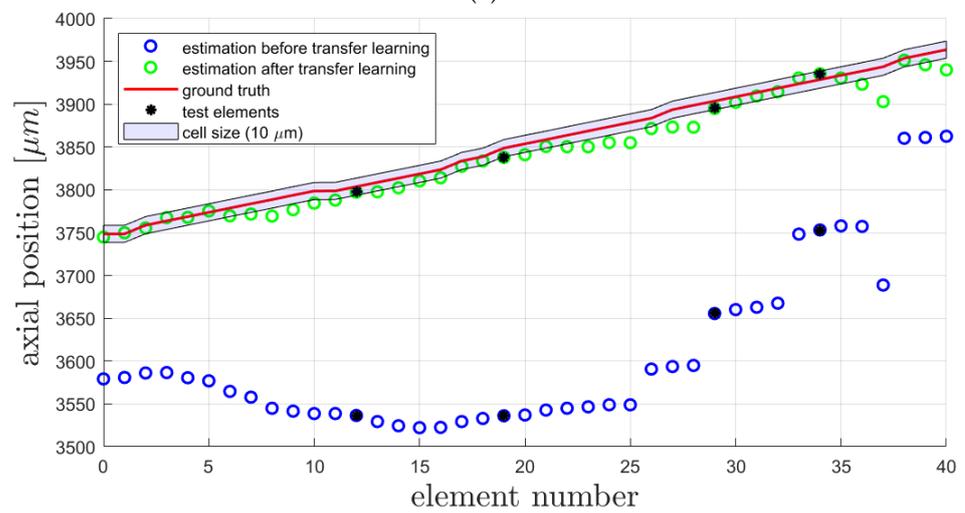
For measurement of the fluorescent light with the detector, a longpass filter with the cut-on wavelength 550 nm (Thorlabs FEL0550) was placed between lens L1 and the upper beam splitter. Furthermore, a voltage amplifier (Femto DLCPA-200) and another photosensor (Hamatsu H10720-20) were necessary to improve the detection, filter the noise, and detect the fluorescent signal.

The fluorescent beads are detected by performing lateral scans of the above introduced sample, the resulting lateral scans are shown in Figure 6b. The lateral scans were performed using a movable stage actuated by a Newport XPS system (for details see Section 2). Thereby, the y-axis was used as the slow axis, while the x-axis was used as the fast axis. An axial shift in the z-axis was only achieved due to tuning the voltage applied to the AL. The scans originate from different axial positions above and underneath the glass plate. Simultaneously, the lateral scans images from the illumination beam were captured. These images serve as input images for the trained CNN.

The estimations made by the CNN for the in-focus particles are then subtracted to obtain the thickness of the plate in the sample. In this case, the difference of the estimations is $169 \mu\text{m}$, and the ground truth optical pathway is $170 \pm 5 \mu\text{m}$. So, the length through the glass plate with fluorescent particles on top of the mirror and the glass was estimated properly.



(a)



(b)

Figure 5. Comparison of ground truth and estimation of the CNN, here the elements from the test datasets are marked in black: (a) Results after the first training on the first dataset; (b) Results before and after transfer learning. All results are obtained on the second dataset.

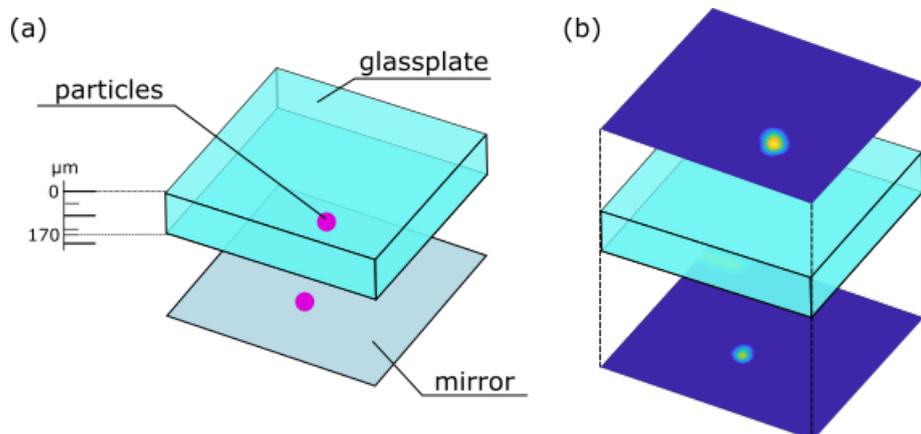


Figure 6. (a) Scheme of the prepared sample consisting of a glass plate covered by fluorescent beads placed onto a mirror. (b) Lateral scans of fluorescent beads above and underneath the glass plate.

5. Conclusions

We presented a modified implementation of the AlexNet architecture for a neural network to estimate the position of the focal plane in an adaptive confocal microscope based on an image of the illumination beam. With this network, it is possible to perform accurate single shot estimation of the focal spot position with an *MAE* under 10 μm , which is the average size of thyroid cells in zebrafish embryos. In previous studies, zebrafish thyroids have already been investigated by adaptive lens scanning.

As a practical application of the network, we reported a thickness measurement of a sample with fluorescent beads, which was performed by our network and the adaptive lens replacing a movable stage. One of the main advantages here is the independence of samples. This is possible because the neural network takes an image from the illumination beam before surpassing the sample as input. So no information from the sample is included in the image for the CNN. Thus, the type of sample is irrelevant for the CNN and makes it quite versatile. However, the network parameters and architecture might have to be adapted for data from other experimental setups. The challenge here is to find the optimal training parameters and grade of data augmentation. However, several rounds of transfer learning should enhance the robustness and generalization ability of the CNN. So, the fine tuning of the training parameters might become less crucial.

Although we had to challenge a small dataset for training and measurement uncertainties, it was possible to develop and train a CNN for single shot focal point estimation in a confocal setup. With this work, we placed a cornerstone for smart microscopy and made a step forward to fast self calibration for adaptive optics.

Author Contributions: J.W.C. and N.K. contributed the concept of adaptive optics for microscopy. K.S. and N.K. designed and built the experimental setup. The network architecture development and training process was done by K.S. and N.K. and J.W.C. supervised the whole research work. K.S. wrote the manuscript. N.K. and J.W.C. revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by DFG grant number CZ 55/32-2.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon request.

Acknowledgments: Special thanks to the group of Ulrike Wallrabe, Laboratory for Microactuators, Department of Microsystems Engineering (IMTEK) at University of Freiburg for manufacturing and providing the adaptive lens for this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Paddock, S.W. Principles and practices of laser scanning confocal microscopy. *Mol. Biotechnol.* **2000**, *16*, 127–149. [[CrossRef](#)]
2. Jabbour, J.M.; Malik, B.H.; Olsovsky, C.; Cuenca, R.; Cheng, S.; Jo, J.A.; Cheng, Y.S.L.; Wright, J.M.; Maitland, K.C. Optical axial scanning in confocal microscopy using an electrically tunable lens. *Biomed. Opt. Express* **2014**, *5*, 645–652. [[CrossRef](#)]
3. Dixon, A.E.; Damaskinos, S.; Atkinson, M.R. Transmission and double-reflection scanning stage confocal microscope. *Scanning* **1991**, *13*, 299–306. [[CrossRef](#)]
4. Hamann, G. Optical 3D Surface Measuring Technology. *Opt. Photonik* **2015**, *10*, 49–51. [[CrossRef](#)]
5. Yasuno, Y.; Makita, S.; Yatagai, T.; Wiesendanger, T.F.; Ruprecht, A.K.; Tiziani, H.J. Non-mechanically-axial-scanning confocal microscope using adaptive mirror switching. *Opt. Express* **2003**, *11*, 54–60. [[CrossRef](#)] [[PubMed](#)]
6. Radner, H.; Stange, J.; Büttner, L.; Czarske, J. Field-Programmable System-on-Chip-Based Control System for Real-Time Distortion Correction in Optical Imaging. *IEEE Trans. Ind. Electron.* **2021**, *68*, 3370–3379. [[CrossRef](#)]
7. Radner, H.; Büttner, L.; Czarske, J. Interferometric velocity measurements through a fluctuating phase boundary using two Fresnel guide stars. *Opt. Lett.* **2015**, *40*, 3766–3769. [[CrossRef](#)] [[PubMed](#)]
8. Koukourakis, N.; Finkeldey, M.; Stürmer, M.; Leithold, C.; Gerhardt, N.C.; Hofmann, M.R.; Wallrabe, U.; Czarske, J.W.; Fischer, A. Axial scanning in confocal microscopy employing adaptive lenses (CAL). *Opt. Express* **2014**, *22*, 6025–6039. [[CrossRef](#)]

9. Förster, E.; Stürmer, M.; Wallrabe, U.; Korvink, J.; Brunner, R. Bio-inspired variable imaging system simplified to the essentials: Modelling accommodation and gaze movement. *Opt. Express* **2015**, *23*, 929–942. [[CrossRef](#)]
10. Schneider, F.; Draheim, J.; Kamberger, R.; Waibel, P.; Wallrabe, U. Optical characterization of adaptive fluidic silicone-membrane lenses. *Opt. Express* **2009**, *17*, 11813–11821. [[CrossRef](#)]
11. Duocastella, M.; Vicidomini, G.; Diaspro, A. Simultaneous multiplane confocal microscopy using acoustic tunable lenses. *Opt. Express* **2014**, *22*, 19293–19301. [[CrossRef](#)] [[PubMed](#)]
12. Ren, H.; Fox, D.W.; Wu, B.; Wu, S.T. Liquid crystal lens with large focal length tunability and low operating voltage. *Opt. Express* **2007**, *15*, 11328–11335. [[CrossRef](#)] [[PubMed](#)]
13. Kopp, D.; Zappe, H. Tubular Focus-Tunable Fluidic Lens Based on Structured Polyimide Foils. *IEEE Photonics Technol. Lett.* **2016**, *28*, 597–600. [[CrossRef](#)]
14. Philipp, K.; Lemke, F.; Scholz, S.; Wallrabe, U.; Wapler, M.C.; Koukourakis, N.; Czarske, J.W. Diffraction-limited axial scanning in thick biological tissue with an aberration-correcting adaptive lens. *Sci. Rep.* **2019**, *9*, 9532. [[CrossRef](#)]
15. Philipp, K.; Smolarski, A.; Koukourakis, N.; Fischer, A.; Stürmer, M.; Wallrabe, U.; Czarske, J.W. Volumetric HiLo microscopy employing an electrically tunable lens. *Opt. Express* **2016**, *24*, 15029–15041. [[CrossRef](#)]
16. Fox, I.; Lee, J.; Pop-Busui, R.; Wiens, J. Deep Reinforcement Learning for Closed-Loop Blood Glucose Control. *Proc. Mach. Learn. Res.* **2020**, *126*, 508–536.
17. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016. Available online: <http://www.deeplearningbook.org> (accessed on 14 January 2021).
18. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]
19. Chen, Q.; Wu, H.; Wada, T. Camera Calibration with Two Arbitrary Coplanar Circles. In Proceedings of the Computer Vision—ECCV 2004, Prague, Czech Republic, 11–14 May 2004; Pajdla, T., Matas, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 521–532.
20. Deutscher, J.; Isard, M.; MacCormick, J. Automatic Camera Calibration from a Single Manhattan Image. In Proceedings of the Computer Vision—ECCV 2002, Copenhagen, Denmark, 28–31 May 2002; Heyden, A., Sparr, G., Nielsen, M., Johansen, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2002; pp. 175–188.
21. Bujnak, M.; Kukulova, Z.; Pajdla, T. Robust Focal Length Estimation by Voting in Multi-view Scene Reconstruction. In Proceedings of the Computer Vision—ACCV 2009, Xi'an, China, 23–27 September 2009; Zha, H., Taniguchi, R.I., Maybank, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; pp. 13–24.
22. Workman, S.; Greenwell, C.; Zhai, M.; Baltenberger, R.; Jacobs, N. DEEPPFOCAL: A method for direct focal length estimation. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015; pp. 1369–1373. [[CrossRef](#)]
23. López, M.; Marí, R.; Gargallo, P.; Kuang, Y.; Gonzalez-Jimenez, J.; Haro, G. Deep Single Image Camera Calibration with Radial Distortion. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 11809–11817. [[CrossRef](#)]
24. Yan, H.; Zhang, Y.; Zhang, S.; Zhao, S.; Zhang, L. Focal length estimation guided with object distribution on FocaLens dataset. *J. Electron. Imaging* **2017**, *26*, 033018. [[CrossRef](#)]
25. Pinkard, H.; Phillips, Z.; Babakhani, A.; Fletcher, D.A.; Waller, L. Deep learning for single-shot autofocus microscopy. *Optica* **2019**, *6*, 794–797. [[CrossRef](#)]
26. Philipp, K.; Czarske, J. Axial scanning employing tunable lenses: Fourier optics based system design. *OSA Contin.* **2019**, *2*, 1318–1327. [[CrossRef](#)]
27. Philipp, K. Investigation of Aberration Correction and Axial Scanning in Microscopy Employing Adaptive Lenses. Ph.D. Thesis, TU Dresden, Dresden, Germany, 2019.
28. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
29. Falcon, W. PyTorch Lightning. *GitHub Note*. 2019; Volume 3. Available online: <https://github.com/PyTorchLightning/pytorch-lightning> (accessed on 10 March 2021).
30. Jung, A.B.; Wada, K.; Crall, J.; Tanaka, S.; Graving, J.; Reinders, C.; Yadav, S.; Banerjee, J.; Vecsei, G.; Kraft, A.; et al. *Imgaug*. 2020. Available online: <https://github.com/aleju/imgaug> (accessed on 1 February 2020).
31. Newport Corporation. *Motorized Actuator, LTA-HS Integrated with CONEX-CC Controller*; Newport Corporation: Irvine, CA, USA, 2018.
32. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
33. Pawley, J. *Handbook of Biological Confocal Microscopy*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2006.