Autofocus Vision System Enhancement for UAVs via Autoencoder Generative Algorithm

Anwer Ahmed Saleh
College of Computer Science & Information Technology, University of Anbar, Ramadi, Iraq
anw21c1012@uoanbar.edu.iq (corresponding author)

Rabah Nori Farhan
Control Systems and Communications Department, Renewable Energy Research Centre, University of Anbar, Ramadi, Iraq
rabahalobadiy@uoanbar.edu.iq

Received: 25 July 2024 | Revised: 5 September 2024 | Accepted: 8 September 2024

 DOI: https://doi.org/10.48084/etasr.8519

ABSTRACT

The Autofocus (AF) technology has become well-known over the past four decades. When attached to a camera, it eliminates the need to manually focus by giving the viewer a perfectly focused image in a matter of seconds. Modern AF systems are needed to achieve high-resolution images with optimal focus, and AF has become very important for many fields, possessing advantages such as high efficiency and autonomously interacting with Fenvironmental conditions. The proposed AF vision system for Unmanned Aerial Vehicle (UAV) navigation uses an autoencoder technique to extract important features from images. The system's function is to monitor and control the focus of a camera mounted to a drone. On an AF dataset, the proposed autoencoder model exhibited an amazing 95% F-measure and 90% accuracy, so it can be considered a robust option for achieving precision and clarity in varying conditions since it can effectively identify features.

Keywords-Unmanned Aerial Vehicle (UAV) navigation; autofocus; feature extraction; autoencoder

I. INTRODUCTION

Autofocus (AF) is a major feature of high-resolution digital cameras [1]. Compared to human-focused attempts, AF systems are faster and more accurate. Researchers and engineers are constantly working on ways to enhance and perfect the currently available AF systems [2]. Video cameras employ the AF approach, which involves altering the focusing lens, to improve the high-frequency elements of an image. Higher frequency components are usually seen in sharply focused images compared to blurry ones within the same scene. By analyzing the overall high-frequency elements of the video signal to identify a frame/field, one can determine the ideal focus point within the focus range. These evaluations are called focal values [3]. To get good shots when shooting from a distance, most cameras use the AF method [4]. Research into AF systems is ongoing in many fields, such as optical microscopes. With smart camera systems and industrial imaging monitoring applications, AF based on pictures is essential for UAV (Unmanned Aerial Vehicle) systems due to this method's novel features, ease of use, and rapid reaction times [5, 6]. The standard components of an AF system are a motor, lens, processor, and control components.

AF technology has been around for an extended period and is now considered technologically and commercially robust. Digital single-lens reflex and compact cameras feature a plethora of fast AF solutions. These cameras may give the viewer perfectly focused images in a flash since they eliminate the need for focus adjustment and offer users more convenience. Military and healthcare are only two of the fields that make extensive use of AF system applications. The development of this system has the promise of reducing workload, improving detection accuracy and efficiency, and saving time [7, 8]. Color information is crucial in AF [9]. The primary objective of this study is to design a more robust and efficient drone system that can improve image quality for various applications by adjusting the camera focus in real-time using deep learning methods, addressing the many problems with current drone AF solutions.

Image quality could be diminished by external factors and device constraints [10]. AF is one of the most important features of imaging devices. Image enhancement is an essential tool in many industries, including photography and medical imaging [11]. The production of out-focused and unclear images is one of the basic problems that occur in UAVs. Finding practical solutions for AF on UAVs has become more important, given their widespread use. To address these problems, a more effective and reliable system must be designed by improving the quality of the images used through real-time adjustment of the camera focus. In [12], a learning-based solution was presented for AF in digital cameras. In [13], digital holography reconstruction research and deep learning techniques such as Convolutional Neural Networks (CNNs) were used to determine distances in multi-sectional objects. CNNs can make faster distance predictions without a physical setup or reconstruction. In [14], the Passive AF Based-Brightness and Contrast (PA Based PC) approach was proposed, which relies on determining the relative brightness of the R, G, and B channels of an RGB image before focusing the camera on the channel with the highest signal intensity. In [15], a new AF method was designed for use with transmission light on the surface structures of transparent materials, employing a two-step optimization scheme. In [16], a single-shot microscopic AF was proposed, which used a single natural image to estimate the defocus distance, improving the model's ability to extract detailed information from images. In [17], a deep learning program was proposed, which can automatically grade the quality of an image based on its input and the current state of control parameters such as contrast, focus, and brightness.

II. RESEARCH METHODOLOGY

The proposed model was trained and tested with the help of the large dataset in [12]. These images were captured from a Dual Pixel (DP) sensor. A collection of images taken at varied focus distances is called a focal stack. Focal stacks, which are composed of many photos shot from different angles and with varying degrees of magnification, represent most of the dataset. Outdoor scenes with varying degrees of illumination were included. Color, texture, scene features, and depth are all very variable in these focal stacks. The images were then produced by applying varying degrees of zoom. Using these images, the autoencoder model was trained and evaluated to find the optimal camera focus. The collection contains images with varying zoom levels, target distances, lighting conditions, etc. The images were taken indoors and outdoors and had a wide range of resolutions. An identical type of phone was used to gather five images for each capture. The DP data for these images have a resolution of 1512×2016.

The dataset was preprocessed to resize and augment its images. The AF dataset contains photos with dimensions of 1512×2016, which were resized to 256×256 for the autoencoder model. This size is ideal for obtaining clear images with important details. With respect to zoom, target distance, lighting, and important attributes, each image in the dataset is distinct from the others.

Image augmentation is a method to generate additional AF images from current ones to increase the size of the dataset. The augmented data were created from original images that were altered (rotating=20°, flipping=0.2, and zooming=0.2 from different viewpoints) to make the training set more diverse. By providing new, and different instances for training datasets, augmented data enhances the model's efficiency and accuracy. Data augmentation is a great tool to improve autoencoder models by adding new and unique images to training datasets. Autoencoder models work more effectively and correctly when their datasets are rich in an instance and well-structured. Data augmentation is a powerful method to build efficient and reliable autoencoder models and enhance their accuracy.

The next step was to split the dataset into two parts: one that was used for testing (20%) and the other for training (80%). Training and testing the autoencoder network allowed it to find the optimal position for the camera focus. The next step was to determine the best camera focus parameters by analyzing photos with the autoencoder [18]. Autoencoders combine encoder and decoder training. The input is transferred to the initial encoder layer, where all layers decrease the data dimensions. The output layer captures the most relevant signals, and the decoder layer is successively constructed [19].

A. Using Autoencoder in Autofocus System for UAVs

In both military and civilian sectors, AF technologies have been a boon, since they are essential for rapid and precise target capture. Optical AF mechanisms can be categorized as either active or passive focusing. Active focus approaches incur higher production costs and add technological complexity to the optical system due to the addition of external sensors. Passive focusing, on the other hand, changes focus depending on the picture quality [20]. Two primary methods have been investigated: software and hardware AF. The former uses algorithms applied to the images, while the second utilizes specific hardware elements. Available commercial hardware AF systems often use a confocal displacement setup with white light or a laser to determine the in-focus point. These systems are quick, have high cost, need materials that reflect light well, and produce artifacts. Focusing on less expensive hardware is possible according to some alternative hardware-based strategies that have been suggested. These approaches cannot be implemented without additional hardware, which includes a phase sensor and an LED pattern illuminator [21-23], whereas some systems might not be able to accommodate the installation of such gear.

Many deep learning algorithms have been proposed for AF, and deep learning has recently become prominent as a method for many computing tasks. Applying deep reinforcement learning, a method was proposed in [22] that used a series of fine- and coarse-grained processes to bring the system into focus following the development of an AF policy. Hardware AF systems, common to DSLRs, employ contrast detection to evaluate an image's contrast and specific sensors analyze the phase difference between light rays to find focus. Software-based AF system autoencoder models analyze images using neural networks and predict the best focus adjustment. The models identify patterns and improve focus. The proposed software-based AF system for UAVs utilizes an autoencoder model to address the issues of traditional hardware-based systems without requiring additional hardware components.

Autoencoders are a special kind of artificial neural network that can learn to store images in compressed format. There are two primary components: The encoder compresses the original images, and the decoder rebuilds the original images using the compressed representation. The objective is to understand how to reduce the data size while keeping crucial details and then use this smaller representation to reconstruct the images as close as possible to the original ones with low loss. Image quality can be analyzed with an autoencoder. The encoder reduces the size of the image to help the system determine where the focus needs adjusting. With the autoencoder's ability to differentiate between in-focus and out-of-focus areas, the AF system can zoom in on the critical areas of the image and find the best possible focus points with less effort and time spent compressing the original image.

As UAVs are always on the move, it is difficult to keep the subject and focus continuous in traditional focus cameras, making maintaining focus challenging. The proposed system provides excellent focused images and high accuracy under different environmental conditions, addressing the issues the traditional AF systems face without requiring additional hardware components. The autoencoder was trained using the AF dataset involving in-focus and out-of-focus images. After the training was complete, the AF system was able to adapt to different movement and lighting conditions. The proposed system is lightweight, flexible, and with lower complexity, and can be used in many fields besides UAVs.

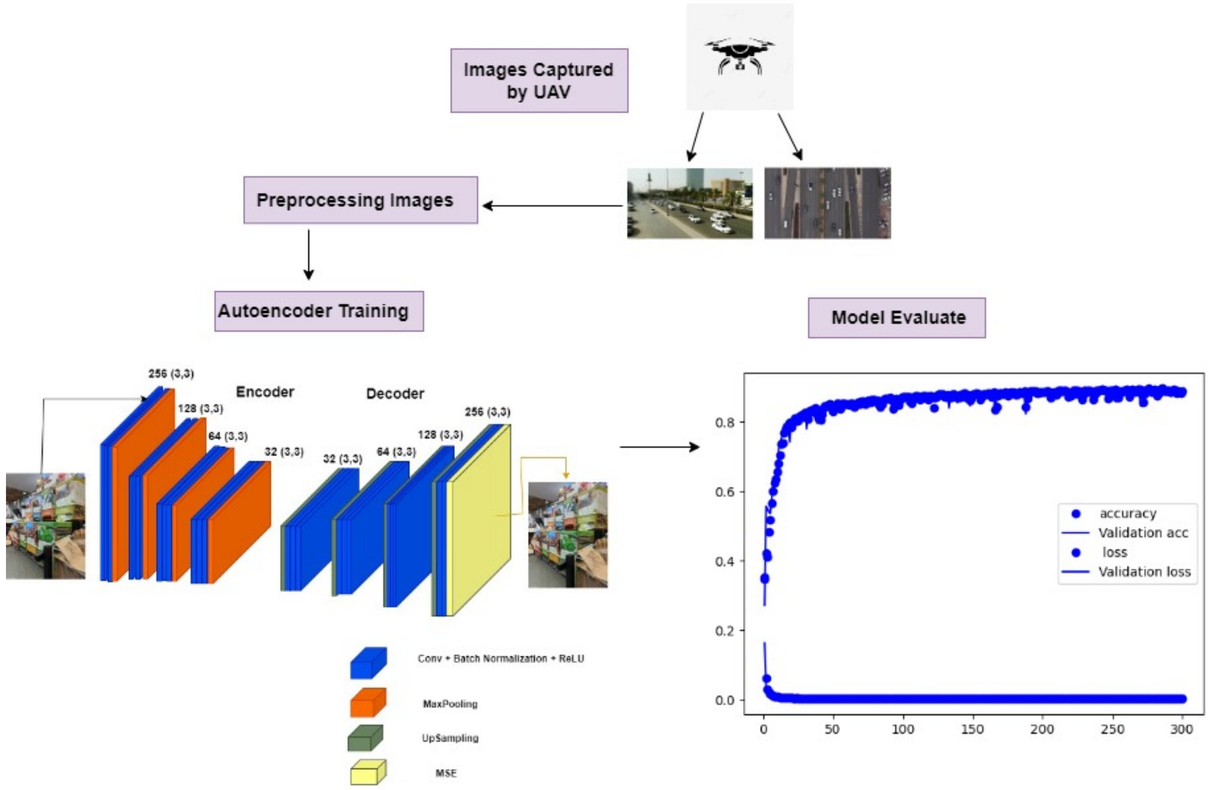

Fig. 1. The proposed AF system.

B. *The Autofocus Vision System for UAV for Image Enhancement*

Visual navigation is an innovative approach to guiding UAVs using computer vision, providing high reliability and autonomy. However, issues such as low light levels or underperforming sensors call for solutions to improve low-light photography, such as the AF vision system that uses an autoencoder to fine-tune focus in real-time.

C. *The Proposed Model*

Artificial intelligence methods are frequently used for image processing applications. Separate convolutional layers, pooling layers, and fully linked layers make up the CNN layer architecture [24]. Autoencoders use hidden neuron layers to learn to compress and rebuild input data. An autoencoder consists of an encoder, a latent space, and a decoder. Encoders compress input data into latent space to generate lower-dimensional representations. The decoder can then use this representation to restore the initial input data. Through the process of learning this compressed representation, the model can learn to capture the elements that are most important from the input information in a lower-dimensional space. Figure 2 shows the proposed autoencoder.

The input of the autoencoder consists of 256×256 images. The encoder is used to compress the dimensions of the input picture into a representation called latent space. From the input image, the encoder extracts features at various levels using an ordered series of convolutional layers. On each layer, a different set of filters is applied to the original image and generates a map of images that highlights certain patterns and structures. Encoders provide compressed latent space representations of the input images. Latent representation captures the most important elements of the input image, and the result is usually a lower-dimensional representation of the input image. The decoder employs a series of deconvolutional layers that incrementally enhance the feature maps toward the eventual goal of generating an output identical in size to the input image. With the use of filters, each layer upsamples the feature maps, and the result is a reconstructed image.

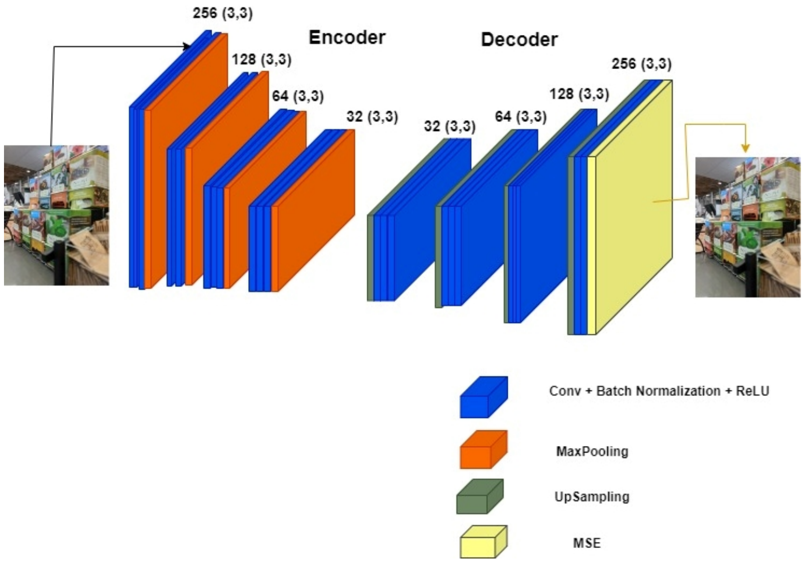

Fig. 2. The proposed autoencoder model.

The initial stage in implementing the proposed method is to collect the dataset and then preprocess it, reshaping the images to 256×256 pixels and augmenting it [19, 25]. The trained model is then created by the autoencoder. The last step is to evaluate the proposed system. The Mean Squared Error (MSE) is used as the training loss function, and the Adam optimizer updated the network weights by backpropagation. The proposed encoder included four conventional layers with Batch Normalization (BN) and the RELU activation function. Max pooling layers with 2×2 filter size were also used, as shown in Figure 2. In the decoder, a 3×3 filter size was used for the 4 conventional layers along with BN and the RELU activation function. A batch size of 32 was chosen and the autoencoder was trained from scratch with 0.2 dropout rate. Training and testing converged to stable values, and no overfitting was observed at 300 epochs.

```
Main Autoencoder Algorithm [26]
Training (b, e, x, l, ϴ)
x= [x1, x2, x3, … x n] is the input
matrix.
e is the number of epochs
is the number of batches
is the learning rate
ϴ = [ W, W^, b^, b] are the parameters
of the autoencoder network
for 0 to e do
  for 0 to b do
    x= φ (W x +b).
    y=φ (W^ y+ b^) where W is the
    weight matrix and b is bias
    vector.
    compute loss which is defined in
    MSE=1/N ∑_{i=1}^N ( x_i - x^_i)^2
    compute the gradient of loss with
    respect to ϴ
    for ϴi, gi in (ϴ, g) do
      ϴ_{i+1}= ϴ_i-l*g_i
    end for
  end for
end for
```

D. Evaluation

Recall, accuracy, precision, and F-score are common measures used to evaluate model performance. These measurements determine the model's performance in different parts [27]. TP, FP, TN, and FN are the counts of True Positive, False Positive, True Negative, and False Negative predictions, respectively. Precision represents the TP results out of the total predicted positives:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall measures how well the model detects the positive cases:

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Accuracy measures the proportion of correct predictions to the total number of predictions, evaluating the model's overall efficiency.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (3)$$

F-measure provides a comprehensive metric that establishes a suitable compromise between precision and recall.

$$F-measure = \frac{2*(Recall* Precision)}{Precision+Recall} \quad (4)$$

Root Mean Square Error (RMSE) denotes an average difference between the model's predicted and actual values [6].

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(xi - x\hat{}i)2} \quad (5)$$

The proposed model achieved 90% accuracy, 0.98 precision, 0.91 recall, and 0.95 F-score. These results show that the model was successful in identifying patterns and was able to make correct predictions.

III. RESULTS AND DISCUSSION

A. Reconstructed Results for the Autoencoder Model

The results show that the autoencoder model can reconstruct AF images from a compression representation close to the original images. The proposed autoencoder model can understand the in-focus and out-of-focus areas. The encoder extracts the features, while the decoder generates the depth maps from these features. The autoencoder model keeps important details and colors in the reconstructed autofocused images. The primary distinction between this approach and previous methods is that it enables the system to learn and analyze in-focus and out-of-focus images, including object distance and focal length, to determine the internal laws and intricate relationships between them. The network will attempt to approximate the association between them as closely as possible. The autoencoder employs object distance, lighting, picture angles, and other information to train the input-to-output relationship in AF. Once the training is finished, the neural network can quickly rebuild the necessary information when it receives new data. The system then uses the trained model to finish its AF of the optical system. This procedure can be carried out without the need for user input, parameter adjustments, or iterations. Compared to traditional previous studies, the focusing speed and precision are greatly improved and the procedure is now more automated and intelligent.

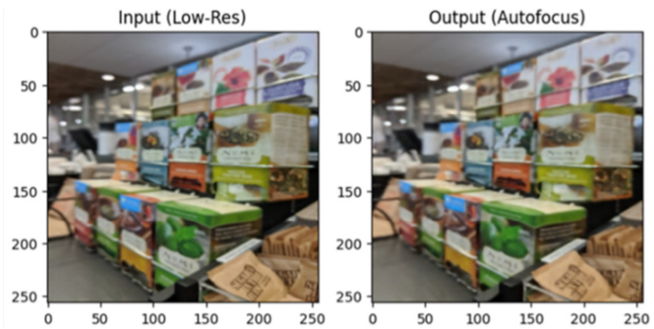

Fig. 3. Original AF images (left), and AF images (right) generated by the proposed autoencoder model.

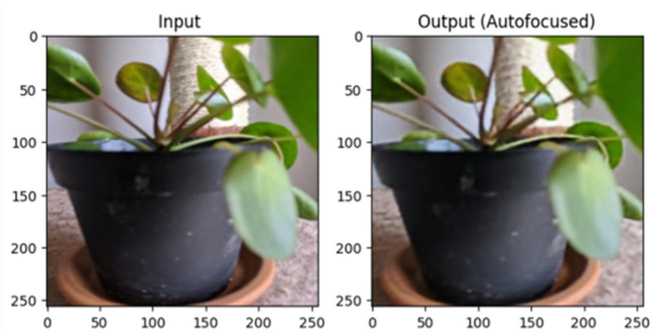

Fig. 4. Original AF images (left) and AF images (right) generated by the proposed autoencoder model.

Two experimental runs were conducted. The network was trained twice (case 1 and case 2) with the same training-testing ratio from the considered dataset. Table I shows the results for the AF accuracy utilizing the proposed autoencoder model in case 1. The accuracy in case 2 was 89%, and compared to the KNN method of [13] (different dataset), the proposed autoencoder model achieved better accuracy. Table II shows a comparison with relevant studies. The proposed autoencoder model outperformed the other approaches with an RMSE of 0.016. The VGG16 approach for AF state definition was proposed in [15] using a different dataset. This technique evaluated and determined focus states by looking at several distinct components, achieving an F-measure of 90.3. The passive AF method [14], using a different dataset, exhibited an accuracy of 80.5%. The AF approach for grayscale images based on deep learning in [20] exhibited an RMSE of 0.201 on a different dataset. The learning-based solution for AF in digital cameras [12] achieved an RMSE rate of 2.446 on the same dataset. MobileNetV1, MobileNetV2, and EfficientNet [16] achieved RMSE rates of 2.92, 1.64, and 3.00, respectively, on a different dataset. VGG19, VGG13, and PACF [17] acquired RMSE values of 1.3250, 1.2035, and 1.2256, respectively. The experimental results show that incorporating an autoencoder in an AF system improves the image quality more than the conventional methods, due to its ability to learn important patterns and extract fine features and details, accelerating AF and reducing processing time. Additionally, the model can handle various scenarios effectively. The proposed method automatically learns and analyzes the extraction features, and the system can learn and update when new images are considered. The system achieved high accuracy and excellent focus under different conditions and can be used in many fields. The use of an autoencoder in an AF system for UAVs represents a significant advancement in visual performance.

TABLE I. COMPARISON BETWEEN THE PROPOSED MODEL AND THE KNN METHOD [13]

	Accuracy	Precision	Recall	F-Score
Case1	0.90	0.98	0.91	0.95
Case2	0.89	0.97	0.96	0.96
Test KNN [13] / validation	0.640 0.627	0.348 0.370	0.151 0.092	0.211 0.147

TABLE II. COMPARISON OF THE PROPOSED SYSTEM WITH RELEVANT STUDIES

Reference	F-score	Accuracy	RMSE
[15]	90.3%	--	--
[14]	--	80.5%	--
[20]	--	--	0.201
[12]	--	--	2.446
MobileNetV1 [16]	--	--	2.92
MobileNetV2 [16]	--	--	1.64
EfficientNet1 [16]	--	--	3.00
VGG19 [17]	--	--	1.3250
VGG13 [17]	--	--	1.2035
PACF [17]	--	--	1.2256
Proposed	0.95%	0.90%	0.016

IV. CONCLUSION

This study presents an AF method as a software system that regulates the focus of a UAV camera to improve image quality for navigation. This system uses an autoencoder to assess photos and find appropriate adjustments for the camera's focus. The main objective of this study is an improved and reliable navigation system tailored for UAVs. With a higher focus state and excellent accuracy, the test results show that the proposed method can effectively and quickly establish AF in many surroundings, with different lighting and flight conditions. The experimental results showed that this approach achieved high performance with an accuracy of 90%. The proposed autoencoder AF method achieved improved navigation performance, and thus the system can increase accuracy and reliability in different environments for many applications, such as military, navigation, crop monitoring, and photography. Future studies should investigate the remaining challenges related to image quality optimization in extremely complicated situations, based on the encouraging results of this research.

REFERENCES

[1] T. H. Tsai and C. Y. Lin, "A new auto-focus method based on focal window searching and tracking approach for digital camera," in *2008 3rd International Symposium on Communications, Control and Signal Processing*, St. Julian's, Malta, Mar. 2008, pp. 650–653, https://doi.org/10.1109/ISCCSP.2008.4537305.

[2] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston Jr., "Comparison of autofocus methods for automated microscopy," *Cytometry*, vol. 12, no. 3, pp. 195–206, 1991, https://doi.org/10.1002/cyto.990120302.

[3] J. S. Lee, Y.-Y. Jung, B. S. Kim, and S. J. Ko, "An advanced video camera system with robust AF, AE, and AWB control," *IEEE Transactions on Consumer Electronics*, vol. 47, no. 3, pp. 694–699, Dec. 2001, https://doi.org/10.1109/30.964165.

[4] Z. Li and K. Wu, "Autofocus system for space cameras," *Optical Engineering*, vol. 44, no. 5, May 2005, Art. no. 053001, https://doi.org/10.1117/1.1901686.

[5] S. Liu, M. Liu, and Z. Yang, "An image auto-focusing algorithm for industrial image measurement," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, Jun. 2016, Art. no. 70, https://doi.org/10.1186/s13634-016-0368-5.

[6] W. Lee, H. S. Nam, Y. G. Kim, Y. J. Kim, J. H. Lee, and H. Yoo, "Robust autofocusing for scanning electron microscopy based on a dual deep learning network," *Scientific Reports*, vol. 11, no. 1, Oct. 2021, Art. no. 20933, https://doi.org/10.1038/s41598-021-00412-5.

[7] H. Zhang and J. Yao, "Automatic Focusing Method of Microscopes Based on Image Processing," *Mathematical Problems in Engineering*, vol. 2021, no. 1, 2021, Art. no. 8243072, https://doi.org/10.1155/2021/8243072.

[8] Y. Zang, L. Liu, W. Gong, H. Yu, W Wang, and C. Zhao, "Autofocus System and Evaluation Methodologies: A Literature Review," *Sensors and Materials*, vol. 30, no. 5, pp. 1165–1174, 2018, https://doi.org/10.18494/SAM.2018.1785.

[9] M. V. Sarode and P. R. Deshmukh, "Image Sequence Denoising with Motion Estimation in Color Image Sequences," *Engineering, Technology & Applied Science Research*, vol. 1, no. 6, pp. 139–143, Dec. 2011, https://doi.org/10.48084/etasr.54.

[10] Y. Said and Y. A. Alsariera, "Hardware Implementation of a Deep Learning-based Model for Image Quality Assessment," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 13815–13821, Jun. 2024, https://doi.org/10.48084/etasr.7194.

[11] S. Chopparapu, G. Chopparapu, and D. Vasagiri, "Enhancing Visual Perception in Real-Time: A Deep Reinforcement Learning Approach to Image Quality Improvement," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14725–14731, Jun. 2024, https://doi.org/10.48084/etasr.7500.

[12] C. Herrmann *et al.*, "Learning to Autofocus," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, Jun. 2020, pp. 2227–2236, https://doi.org/10.1109/CVPR42600.2020.00230.

[13] Z. Ren, Z. Xu, and E. Y. Lam, "Autofocusing in digital holography using deep learning," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XXV*, San Francisco, CA, USA, Feb. 2018, vol. 10499, pp. 157–164, https://doi.org/10.1117/12.2289282.

[14] M. Selek, "A New Autofocusing Method Based on Brightness and Contrast for Color Cameras," *Advances in Electrical and Computer Engineering*, vol. 16, no. 4, pp. 39–45, 2016.

[15] Y. Liao, Y. Xiong, and Y. Yang, "An Auto-Focus Method of Microscope for the Surface Structure of Transparent Materials under Transmission Illumination," *Sensors*, vol. 21, no. 7, Jan. 2021, Art. no. 2487, https://doi.org/10.3390/s21072487.

[16] Z. Hua, X. Zhang, D. Tu, J. Chen, and M. Qian, "Single-shot microscopic autofocus focusing on image detail features," in *International Conference on Optical and Photonic Engineering (icOPEN 2023)*, Singapore, Feb. 2024, Art. no. 17, https://doi.org/10.1117/12.3021718.

[17] H. Kim *et al.*, "Deep-Learning Based Autofocus Score Prediction of Scanning Electron Microscope," *Microscopy and Microanalysis*, vol. 25, no. S2, pp. 182–183, Aug. 2019, https://doi.org/10.1017/S1431927619001648.

[18] M. Sewak, S. K. Sahay, and H. Rathore, "An Overview of Deep Learning Architecture of Deep Neural Networks and Autoencoders," *Journal of Computational and Theoretical Nanoscience*, vol. 17, no. 1, pp. 182–188, Jan. 2020, https://doi.org/10.1166/jctn.2020.8648.

[19] K. Chaitanya *et al.*, "Semi-supervised task-driven data augmentation for medical image segmentation," *Medical Image Analysis*, vol. 68, Feb. 2021, Art. no. 101934, https://doi.org/10.1016/j.media.2020.101934.

[20] Y. Wang, C. Wu, Y. Gao, and H. Liu, "Deep Learning-Based Dynamic Region of Interest Autofocus Method for Grayscale Image," *Sensors*, vol. 24, no. 13, Jul. 2024, Art. no. 4336, https://doi.org/10.3390/s24134336.

[21] H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, "Deep learning for single-shot autofocus microscopy," *Optica*, vol. 6, no. 6, pp. 794–797, Jun. 2019, https://doi.org/10.1364/OPTICA.6.000794.

[22] A. Anikina, O. Y. Rogov, and D. V. Dylov, "Detect to Focus: Latent-Space Autofocusing System With Decentralized Hierarchical Multi-Agent Reinforcement Learning," *IEEE Access*, vol. 11, pp. 85214–85223, 2023, https://doi.org/10.1109/ACCESS.2023.3303844.

[23] T. Ivanov, A. Kumar, D. Sharoukhov, F. Ortega, and M. Putman, "DeepFocus: a deep learning model for focusing microscope systems," in *Applications of Machine Learning 2020*, Aug. 2020, vol. 11511, Art. no. 1151103, https://doi.org/10.1117/12.2568990.

[24] N. I. Sino, R. N. Farhan, and M. E. Seno, "Review of Deep Learning Algorithms in Computational biochemistry," *Journal of Physics: Conference Series*, vol. 1804, no. 1, Oct. 2021, Art. no. 012135, https://doi.org/10.1088/1742-6596/1804/1/012135.

[25] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, "Data Augmentation for Brain-Tumor Segmentation: A Review," *Frontiers in Computational Neuroscience*, vol. 13, Dec. 2019, https://doi.org/10.3389/fncom.2019.00083.

[26] C. Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-Supervised Nets," in *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, Feb. 2015, pp. 562–570.

[27] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning - ICML '08*, Helsinki, Finland, 2008, pp. 1096–1103, https://doi.org/10.1145/1390156.1390294.