

UDC 004.8; 004.93; 535.24
EDN: HJYGLB

PACS: 07.05.Mh; 42.30.Va; 42.62.Eh

Neural network-based methods for optical system measurement

Sh. I. Zaripov^{1,2*}, A. V. Polesskiy¹, N. A. Semenchenko¹, and O. A. Medvedev^{1,3}

¹ RD&P Center ORION, JSC, Moscow, 111538 Russia

* E-mail: mail@tonyfierro.com

² MIREA – Russian Technological University, Moscow, 119454 Russia

³ Bauman Moscow State Technical University, Moscow, 105005 Russia

Received 21.05.2026; revised 1.06.2026; accepted 5.06.2026

The paper presents theoretical research and the development results of an intelligent system for the automated inspection of optical system parameters, including focal length, resolving power, and distortion, across various spectral ranges. The system is based on the YOLO convolutional neural network architecture and a specialized hardware-software complex. By employing a neural network-based approach, the software complex performs measurements of optical parameters without continuous operator intervention. This makes the system highly promising for the mass production of optical devices, as it reduces time of measurement, decreases labor costs, and eliminates the impact of the human factor.

Keywords: convolutional neural network, YOLO, computer vision system, object detection, optics, lenses, automated measurements, focal length, distortion, resolving power.

DOI: 10.51368/2949-561X-2026-3-75-81

For citation: Zaripov Sh. I., Polesskiy A. V., Semenchenko N. A., and Medvedev O. A., Applied Physics, No. 3, 75–81 (2026).

Introduction

The production of optical systems (hereinafter referred to as OS) includes the control of the final products' output parameters, which include focal length, distortion, and resolving power, which are key for characterizing OS for operation as part of optoelectronic devices [1].

The goal of introducing artificial intelligence systems into the technological process of manufacturing lenses for various machine vision systems is to expand domestic production while reducing product costs. This can be achieved through the development of automated systems that enable faster and more accurate measurements of optical parameters with minimal human intervention.

Russia's modern metrological base used in the production of OS includes methods for

controlling parameters; however, it lacks automation, which gives rise to several problems: a long production cycle, low labor productivity in serial production, and subjectivity in measurements.

To carry out automated measurements of focal length, distortion, and resolving power using AI methods, it was necessary to solve the following tasks:

1. Development of auto-positioning and auto-focusing algorithms for OS;

2. Development of a neural network for recognizing the test object's control points, which is to be used for measuring the focal length and distortion of the OS;

3. Development of a neural network for recognizing the test object's groups of lines, which is required to determine the resolving power of the OS;

4. Development of algorithms for analysis and the calculations required to determine the values of focal length, resolving power, and distortion of the OS.

Technical Implementation

The starting point for the development of a control system using artificial intelligence methods was a set of measurement techniques and the technical implementation of a precision measurement setup designed to measure the focal length, distortion, and resolving power in two spectral ranges: 0.4–2.0 μm and 8–14 μm . A general measurement scheme is shown in Figure 1.

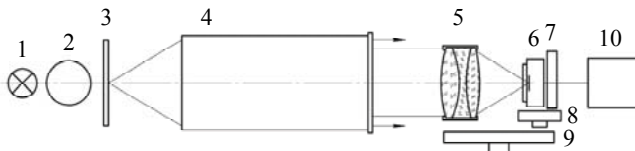


Fig. 1. General scheme for measuring the focal length, distortion, and resolving power of an OS: 1 – radiation source; 2 – integrating sphere (for the spectral range 0.4–2.0 μm); 3 – test object; 4 – collimator; 5 – lens under test; 6 – focal-plane array (FPA); 7 – two-axis linear translator; 8 – linear translator; 9 – rotary platform; 10 – PC with software

The spectral range of the setup is determined primarily by the type of radiation source and FPA used. Radiation source 1 comprises a blackbody model for the spectral range of 8–14 μm and an illumination unit with an incandescent lamp for the spectral range of 0.4–2.0 μm . During operation in the 0.4–2.0 μm spectral range, an integrating sphere 2 may be used if necessary to ensure a uniform radiation distribution in the test object plane. Alternatively, an illumination unit design performing a similar function may be used instead of the integrating sphere. As photodetector array 6 for recording the image in the focal plane of the test object, an uncooled microbolometer and a SWIR camera, which also operates in the visible range, are used alternately as FPA to control lenses in the respective spectral ranges. The test objects 3 are a set of specially developed resolution targets that include groups of lines for determining the resolving power and reference points for estimating the focal length and distortion. To

form a parallel beam of rays simulating an object at infinity, collimator 4 is used. To enable resolving power measurements, the technical implementation of the setup includes an auxiliary projection optical system, which is necessary to address the pixel pitch vs. line images commensurability issue.

Measurement techniques based on the following principles were chosen as objects of interest for automation using neural network methods:

- The method for measuring the lens focal length is based on the classical magnification method. The linear magnification of the optical system consisting of the collimator lens and the lens under test is simultaneously the ratio of the object size to its image size and the ratio of their focal lengths. Therefore, the lens focal length $f'_{\text{к.о}}$ is determined by one of the following formulas (depending on whether the collimator length is known with sufficient accuracy) or by measuring the angular dimensions of the test objects using a theodolite:

$$f'_{\text{к.о}} = f'_{\text{колл}} \cdot \frac{y'}{y}, \quad f'_{\text{к.о}} = \frac{y'}{\text{tg}\varphi}, \quad (1)$$

where $f'_{\text{колл}}$ – focal length of the collimator lens, mm; y – linear size of the test object, mm; y' – linear size of the image of the test object, mm; φ – angular size of the test object, degrees;

- The method for measuring the resolving power of lenses is based on determining the maximum resolvable line target through the lens under test, which corresponds to the smallest resolvable distance between the centers of two adjacent bright and dark lines. The desired resolving power limit $R_{\text{к.о}}$, lines/mm, is determined based on the angular size of the element of the maximally resolvable target $\overline{\varphi}$, degrees, using the formula:

$$R_{\text{к.о}} = \frac{1}{f'_{\text{к.о}} \cdot \text{tg}\overline{\varphi}}; \quad (2)$$

- The method for measuring lens distortion is based on measuring the deviation of the image point coordinates from the calculated value when the lens is rotated using rotary platform 9. This deviation corresponds to the shift of the object

image position at the edge of the lens field caused by the OS distortion. The distortion value v , %, is calculated by the formula

$$v = \frac{y' - l}{l} \cdot 100 \%, \quad (3)$$

where $l = f'_{n.o.} \cdot \tan \omega$ – calculated displacement of the test object image, mm, when the lens is rotated by angle ω , deg; y' – actual displacement of the test object image, mm.

The system is controlled via the interface of the developed software, which implements communication protocols with stepper motor controllers and real-time video stream capture.

Automatic Positioning and Focusing

The basic functions of the automated system are the ability to automatically focus and align the test object's center with that of the FPA. These processes occur during preparation for optical parameter measurements on the bench as follows: automatic positioning of the test object image at the center of the FPA is performed using a two-axis linear translator 7 (see Fig. 1), followed by automatic focusing using linear translator 8.

System focusing is defined as aligning the focal plane of the lens under test with the object plane of the auxiliary projection optical system and is essential for the correct operation of the system and for improving the accuracy of the measurements performed. To automate this process, the system employs a sharpness estimation method based on the Tenengrad criterion [2], which relies on the analysis of image brightness gradients. The idea of the method is as follows: the sharper the image, the stronger the brightness transitions (object edges) and, consequently, the higher the gradient values. This method is based on the Sobel operator and demonstrates high noise immunity in IR frames with elevated noise levels [3]. Accordingly, the autofocus algorithm is reduced to scanning the range of the linear translator responsible for focusing with a predefined step. At each stop, a frame is captured and the sharpness value is calculated using the Tenengrad method. The highest value is saved in the system, and after scanning the entire range, the linear translator is

set to the position with the maximum gradient value.

System positioning involves two stages: axis calibration, which involves calculating the number of motor steps required for the linear translators to move 1 mm, and shifting along each axis to the central point of the image. Axes are calibrated using an alignment table – a set of adjustments for fine-tuning the position. The system uses a neural network to recognize the test object's center and saves the initial position, then the axes are shifted by 1 mm and the recognition is repeated to determine the final position; then the difference between the two positions is calculated, i.e. the distance in pixels traveled by the system during a physical shift of the axis by 1 mm. Several iterations of the shift by a maximum of 1 mm are performed to ensure a more precise calibration. Once the physical distance required to bring the test object's center to the center of the image matrix is determined, the system shifts the axes.

Development of Neural Network Models

Due to recent progress in the field of neural networks, they can be used even with limited computing resources [4, 5], which undoubtedly offers potential for their multifaceted integration into measurement equipment. This opens the way to the development of intelligent systems capable of performing fast and accurate data analysis directly during data acquisition, thus enabling prompt decision-making without significant delays. This increases the operational efficiency of systems under time and resource constraints, while also reducing dependence on external computing infrastructure. Such systems can adapt to changing environmental conditions and ensure a high-level of automation across the entire production process. As a result, their range of applications expands - from single measurements to series production.

Existing classical non-neural network methods are also based on image analysis (including convolution computations), but they are not adapted to multi-stage measurements and require the development of complex algorithms, which is particularly difficult to implement in universal setups designed for measuring various optical systems that differ, for example, in spectral ranges or focal lengths.

With the system requirements taken into account, various families of neural network architectures were analyzed [6–12]. Modern single-stage architectures were considered for solving the problem of recognizing control points and line groups, as they provide the best ratio of speed and accuracy during computations run on a central processing unit without the use of a discrete graphics accelerator. Among them, three main directions stand out: classical detectors, modern universal models, and ultra-lightweight mobile architectures. Transformer models, being the most relevant trend in computer vision, are also distinguished. Despite their high accuracy, they were not considered for this system due to their excessive demands on processor resources.

Under limited performance conditions, such architectures operate significantly slower than their convolutional counterparts. Nevertheless, for completeness of analysis, transformer models have been included in the comparative characteristics of architectures provided in Table 1.

The conducted analysis allowed to establish that modern single-stage architectures of the YOLO family provide the best ratio of localization accuracy for small objects, computational complexity, and ease of integration. In contrast to transformer-based models, they exhibit stable real-time performance on the target hardware, which is why they were selected for the system implementation.

Table 1

Comparative table of single-stage neural network architectures

Architecture	Representatives	Accuracy for small objects	Complexity of integrating in the system	Used in real-time mode
Ultra-lightweight	PicoDet, NanoDet	Medium	Medium	Yes
Classical, single-stage	SSD, RetinaNet	Medium	Low	Limited
Modern universal	YOLO	High	Low	Yes
Transformer	RT-DETR	High	Medium	Limited

For training the neural network models, over 8,000 images were captured in the short-wave visible-to-IR (0.4–2.0 μm) and long-wave (8–14 μm) IR spectral ranges for lenses with focal lengths of 10 mm and 35 mm.

The neural network model was tested during measurements of the developed lenses with known focal lengths using of specially designed resolution targets. The line width of these targets corresponds to the magnification produced by the "lens under test – collimator lens" system, so that the image of the test objects is clearly distinguishable and readable. This is required not only for obtaining reliable measurement results, but also for the stable operation of the neural network model. These targets can also be used for testing lenses with other focal lengths, provided that the image quality is maintained.

The test objects were labeled as follows to train the neural network (see Fig. 2): red – labels for measuring focal length; blue – the center of the line target, used for positioning and for

measuring distortion; green – groups of lines for determining the resolving power limit. The coordinates of the regions of interest are calculated by finding their geometric center.

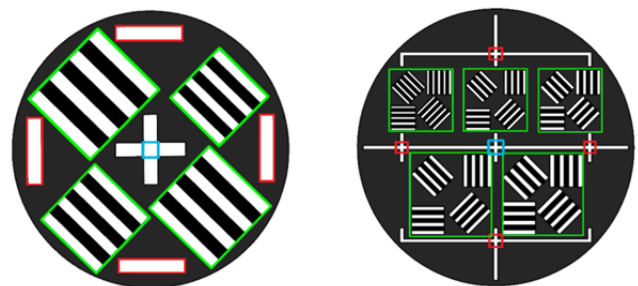


Fig. 2. Labeling of line targets (on the left – for OS with $f' = 10$ mm, on the right – $f' = 35$ mm)

Each group of lines highlighted in green has its own spatial resolution value with a step of 5 lines/mm, corresponding to the focal length of the lens under test, which is set by the operator in the software. The groups on the target in the right part of Fig. 2 are arranged in ascending order, from top to bottom. The line groups on the left

target are arranged in a “crisscross” pattern: the smallest group has the lowest spatial resolution and is followed by a group located diametrically opposite to the first one, the next group is positioned to the left of the second one, and finally, the last group is the largest and, accordingly, has the highest spatial resolution. The results are analyzed and calculated taking into account these specific arrangement features of the groups on the coordinate axes of the image matrix.

To enhance the robustness of the trained neural network models [13], augmentation techniques [14] were applied, including geometric transformations (rotation, scaling) and brightness and contrast adjustments to simulate temperature drift.

Although the neural network models trained on these data successfully prove that they are able to memorize the patterns of the resolution targets, the potential for improving the dataset and developing image preprocessing algorithms to enhance the accuracy and increase the speed of model inference has not been exhausted. In-depth modification of image preprocessing algorithms requires shall be studied individually and was not considered at this stage.

Evaluation of Model Performance

The neural network models were trained using various YOLO modifications ranging from the 8th to the 26th generation, in different configurations: from the most compact and fast (Nano) to the large-scale (Large) versions. Training was performed by varying hyperparameters and environmental conditions. The best results were achieved with medium-sized models: YOLOv8m for the control point detection task, and YOLO26m for recognizing line groups on the resolution targets.

The performance quality of the models was evaluated using a set of metrics that allows evaluating both the correctness of object recognition and the accuracy of their placement in a frame. To interpret these metrics, it is necessary to understand the four possible scenarios of neural network performance:

1. True Positive (TP) – the neural network correctly detected an object where one exists;

2. False Positive (FP) – “false alarm”, the neural network mistook an empty area or noise for an object;

3. False Negative (FN) – a miss of a target, where the neural network failed to detect an existing object;

4. True Negative (TN) – the neural network correctly determined that an empty area contains no object (this parameter is usually not considered in object detection tasks, because the number of empty “areas” in a frame is infinitely large).

Based on these scenarios, the key metrics presented in Table 2 are calculated:

– Precision is the proportion of actual objects among all those that the neural network labeled as detected. High precision means that the system produces very few false alarms. It is calculated by the formula:

$$P = \frac{TP}{TP + FP}; \quad (4)$$

– Recall is the proportion of detected objects out of the total actual number of objects. High recall indicates that the system finds almost all objects and minimizes misses. It is calculated by the formula:

$$R = \frac{TP}{TP + FN}; \quad (5)$$

To evaluate the localization quality, the IoU (Intersection over Union) metric is used, which is the ratio of the intersection area of the predicted and reference frames to their union area (0 = no overlap, 1 = perfect match).

The central performance indicator is mAP (mean Average Precision), which combines the precision and recall metrics. Depending on the strictness of localization requirements (frame overlap), two modifications are distinguished:

– mAP@0.5: calculated under the condition that an object is considered detected (TP) if the overlap area of the predicted and reference frames is at least 50%. This metric characterizes the overall ability of the model to recognize objects;

– mAP@0.5:0.95: represents the average value at different IoU thresholds (from 50% to 95% with 5% increments). This metric is stricter

and more critical in terms of the fitting quality of the bounding box boundaries and allows evaluating the model's robustness to geometric inaccuracies and the precision of object coordinate determination.

Thus, the YOLOv8m model demonstrated high detection quality for control points ($\text{mAP}@0.5 = 0.975$), providing high precision and recall values. This means that in 97.5% of cases, the system correctly finds, classifies, and determines the location of the control point in the

image, allowing only a minimal percentage of errors in the form of misses or false positives. At the same time, the $\text{mAP}@0.5:0.95$ metric was lower, which means lower localization accuracy under strict IoU thresholds. It should be noted that system performance could be further improved by developing and integrating additional algorithms for control point analysis and center finding, but the investigation of this possibility remains beyond the scope of this paper.

Table 2

Results of training neural networks models

Task	Model	Precision	Recall	$\text{mAP}@0.5$	$\text{mAP}@0.5:0.95$
Recognition of test object's control points	YOLOv8m	0.956	0.955	0.975	0.598
Recognition of test object's target lines	YOLO26m	0.813	0.932	0.874	0.737

The YOLO26m model demonstrated lower classification precision when recognizing line groups ($\text{mAP}@0.5 = 0.874$), but it provided better localization stability, as confirmed by a higher $\text{mAP}@0.5:0.95$ value. In other words, the model correctly detects and classifies line groups in 87.4% of cases.

Thus, the combined use of these models effectively addresses the comprehensive task of optical system parameter control. Despite differences in error characteristics, the achieved precision for each subtask provides a sufficient basis for the correct operation of the entire system, ensuring the necessary balance between key point detection and the stability of line group recognition.

Conclusion

The developed automated system for controlling lens parameters, which is based on two neural network models (YOLOv8 m – for measuring focal length and distortion; YOLO26 m – for measuring resolving power), allows controlling optical systems across various spectral ranges (0.4–2.0 and 8–14 μm).

The advantages of this system include speed, precision, and independence from the human factor, making its application particularly

relevant in the mass production of optical systems. The system has demonstrated a reduction in labor efforts during measurements: checking all lens parameters (focal length, distortion, and resolving power) with this intelligent system takes less than 15 minutes, whereas manual measurement by a human operator can require over 60 minutes.

The first neural network model, based on YOLOv8m, detects control points almost perfectly: they are recognized absolutely correctly in 97.5% of cases. This is a high figure, which indicates that errors in point detection are nearly eliminated.

When recognizing the test object's lines, the second model, based on YOLO26m, shows a slightly lower accuracy of 87.4%. This means that when the system encounters a group of lines, it correctly identifies it and accurately isolates it in 87.4% of cases. At the same time, the final measurement error does not exceed the values established by the measurement procedure.

The system was tuned and tested on a computer without a discrete graphics accelerator. All neural network computations were performed using the central processing unit. Despite this, the system enables fast measurements with the required precision, thus confirming the correct selection of models and approaches.

REFERENCES

1. Kreopalova G. V., Lazareva N. L., and Puryaev D. T. Optical Measurements: A Textbook for Universities / Ed. Puryaev D. T. Moscow: Mashinostroenie, 1987.
2. Sun Y., Duthaler S., and Nelson B. J., Microscopy Research and Technique **65** (3), 139–149 (2004).
3. Gashnikov M. V. Methods of Computer Image Processing / Ed. Soifer V. A., 2nd ed. Moscow: FIZMATLIT, 2003.
4. Liu H.-I., et al., ACM Computing Surveys **57** (6), 1–42 (2024). <https://doi.org/10.1145/3657282>.
5. Bouguettaya A., Kechida A., and Taberkit A. M., International Journal of Informatics and Applied Mathematics **2** (2), 28–44 (2019).
6. Redmon J., Divvala S., Girshick R., and Farhadi A. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788 (2016). <https://doi.org/10.1109/CVPR.2016.91>.
7. Yaseen M. What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector, 2024. arXiv preprint. arXiv:2408.15857.
8. Liu W., et al. Computer Vision – ECCV 2016. Lecture Notes in Computer Science **9905**. Cham: Springer, 21–37 (2016). https://doi.org/10.1007/978-3-319-46448-0_2.
9. Mittal P., Artificial Intelligence Review **57**, Art. 242 (2024). <https://doi.org/10.1007/s10462-024-10877-1>.
10. Lin T.-Y., et al. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2980–2988 (2017). <https://doi.org/10.1109/ICCV.2017.324>.
11. Andriyanov N. A., Dementiev V. E., and Tashlinskiy A. G., Journal of Communications Technology and Electronics **67** (10), 1015–1034 (2022).
12. Zhao Y., et al., Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 16965–16974 (2024). arXiv:2304.08069.
13. GOST R 70462.1-2022. Information technology. Artificial intelligence. Assessment of neural network robustness. Part 1. Overview. Moscow: Standartinform, 2022.
14. Emelyanov S. O., Ivanova A. A., Shvets E. A., and Nikolaev D. P., Sensory Systems **32** (3), 236–245 (2018). <https://doi.org/10.1134/S0235009218030058>.
15. Padilla R., Netto S. L., and da Silva E. A. B., International Conference on Systems, Signals and Image Processing (IWSSIP), 237–242 (2020). <https://doi.org/10.1109/IWSSIP48289.2020.9145130>.

About authors

Zaripov Shamil Ilnurovich, programmer engineer, RD&P Center ORION, JSC (9 Kosinskaya Ulitsa, Moscow, 111538, Russia); MIREA – Russian Technological University (78 Prospect Vernadskogo, Moscow, 119454, Russia). E-mail: mail@tonyfierro.com

Polesskiy Alexey Viktorovich, Doctor of Technical Sciences, Director for Research and Development, RD&P Center ORION, JSC (9 Kosinskaya Ulitsa, Moscow, 111538, Russia). E-mail: polesskiyav@orion-ir.ru SPIN code: 3992-4520, Author ID 571156

Semenchenko Natalia Alexandrovna, head of department, RD&P Center ORION, JSC (9 Kosinskaya Ulitsa, Moscow, 111538, Russia). E-mail: nafara@mail.ru SPIN code: 8094-5126, Author ID 775858

Medvedev Oleg Aleksandrovich, RD&P Center ORION, JSC (9 Kosinskaya Ulitsa, Moscow, 111538, Russia); Bauman Moscow State Technical University (Bld. 1, 5 2nd Baumanskaya, Moscow, 105005, Russia). E-mail: om15032006@yandex.ru