ANNOTATION

of the dissertation work by Mutalova Zhazira Satkanovna on the topic "Development of a Mathematical Model for Image Recognition in Video Streams Based on Neural Networks" submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy (PhD) in the educational program 8D06101 – "Big Data Analytics"

The relevance of the research topic is determined by the need to improve mathematical methods and algorithms for image recognition in video streams based on neural networks in order to increase classification accuracy, robustness to external influences, and the efficiency of intelligent computer vision systems.

In the context of the rapid development of artificial intelligence and computer vision technologies, tasks related to the automatic analysis and recognition of images in video streams are becoming increasingly important. Real-time video data processing methods are widely applied in intelligent video surveillance systems, transport security, robotics, medicine, industrial automation, and human behavior monitoring systems.

Despite significant advances in the field of computer vision, existing image recognition methods do not always provide the required accuracy and robustness under conditions of changes in illumination, background, scale, viewpoint, and object dynamics. An additional challenge is the high dimensionality of video data and the need to take into account spatiotemporal characteristics, which requires further development of mathematical models and information processing algorithms.

In this regard, the development of a mathematical model for image recognition in video streams based on neural networks is of both scientific and practical importance, as it enables improved classification accuracy, more efficient extraction of informative features, and increased robustness of the system to external disturbances. The use of modern neural network architectures makes it possible to automate feature representation learning and optimize video data processing in intelligent computer vision systems.

The aim of the doctoral dissertation is to develop a mathematical model for image recognition in video streams based on neural networks and to create a face recognition system based on the proposed models and algorithms.

Tasks of the dissertation research:

– to review and conduct a comparative analysis of domestic and international scientific studies in the field of image recognition in video streams based on neural networks;

– to develop a mathematical model for preprocessing images affected by noise and face recognition in video streams based on convolutional neural networks and Retinex algorithms;

– to develop a multi-stage model based on a convolutional neural network aimed at improving the accuracy of image recognition in video streams and optimizing frame processing time;

– to design a convolutional neural network architecture focused on improving data classification accuracy and processing efficiency;

– to develop a face recognition system for images in video streams based on the proposed mathematical models and algorithms.

Object of the research: processes of image recognition in video streams.

Subject of the research: a mathematical model for image recognition in video streams based on neural networks.

Research methods. To solve the tasks set in the dissertation, a set of theoretical and applied research methods was used. Mathematical modeling methods were effectively applied to construct a mathematical model of the image recognition process in video streams.

For automatic detection and classification of images, the use of convolutional neural networks (CNN) was proposed. The methodological basis of the research includes systems theory, systems analysis, statistical data processing methods, as well as methods of computational experiments and analysis of obtained results.

Statements submitted for defense:

1. A mathematical model for preprocessing images affected by noise and face recognition in video streams based on convolutional neural networks and Retinex algorithms has been developed and substantiated.

2. A multi-stage model based on a convolutional neural network aimed at improving image recognition accuracy in video streams and optimizing frame processing time has been constructed.

3. A convolutional neural network architecture focused on improving data classification accuracy and processing efficiency has been developed.

Main results of the research. As a result of the conducted study, the following key findings were formulated:

A mathematical model for image preprocessing and face recognition based on convolutional neural networks and Retinex algorithms was developed. Within the proposed model, the processes of non-uniform illumination correction, contrast normalization, and noise suppression were formalized, which ensured improved quality of the input data. It was demonstrated that the application of Retinex algorithms at the preprocessing stage increases the informativeness of the feature space and significantly improves face recognition accuracy, especially in complex visual environments with varying illumination conditions.

To improve the accuracy of image recognition in video streams and optimize frame processing time, a multi-stage CNN-based model was developed. The proposed architecture incorporated Faster R-CNN, EfficientNetB0, Self-Attention, and FaceNet models. The Faster R-CNN model provided highly accurate object detection in complex scenes and, when using the ResNet50 FPN backbone, achieved 100% accuracy with a minimum loss value of 0.0102. However, the total training time reached 183 minutes, indicating high computational costs of this architecture. The MobileNetV3 Large architecture demonstrated the highest training speed, reducing the total training time to only 40 minutes. At the same time, the increase in the loss value to 0.1807 during the final epochs indicated reduced model stability. The EfficientNetB0 model combined with the Self-Attention mechanism achieved 100% accuracy as early as the third epoch, while the final loss value reached 0.0289. The Self-Attention mechanism enabled the model to focus on the most informative regions of the image, thereby improving recognition quality under complex and varying illumination conditions. The FaceNet architecture provided reliable face comparison based on Euclidean distance by transforming facial images into 128-dimensional

embeddings. As a result, the proposed multi-stage model ensured highly accurate and stable real-time face recognition in video streams.

To improve data classification accuracy and processing efficiency, a multi-level convolutional neural network architecture was developed. In the proposed Nexiface architecture, the Retinex preprocessing module, Faster R-CNN detector, EfficientNetB0 model, Self-Attention feature extraction module, and FaceNet identification model were integrated into a unified processing pipeline. The Retinex algorithm reduced the influence of non-uniform illumination and improved image quality, thereby increasing the stability of the neural network. The two-stage Faster R-CNN detection architecture provided highly accurate localization of facial regions. Owing to the lightweight EfficientNetB0 architecture, computational costs were reduced while maintaining high processing speed and efficient extraction of significant facial features. The Self-Attention mechanism enabled the model to focus on the most informative regions of the image, such as the eyes, nose, and facial contours, which enhanced the informativeness of the feature space. The FaceNet model generated 128-dimensional embedding vectors and, through the use of triplet loss, ensured the convergence of vectors belonging to the same person and the separation of vectors corresponding to different individuals, thereby improving classification accuracy. The developed architecture was adapted for real-time operation and provided processing of up to 30 frames per second from each camera while maintaining an overall latency below 100 ms.

The practical implementation of the obtained results is confirmed by their deployment in technology companies (in particular, LLP “QUARES”), which enabled the automation of intelligent video analytics processes and improved the reliability of monitoring systems in complex operational conditions. Thus, the set of developed mathematical models and algorithms represents a complete scientific and technological solution to an urgent computer vision problem.

Scientific novelty.

1. A mathematical model for processing large-volume and high-speed video data under big data conditions and face recognition has been developed based on the integration of the Retinex algorithm and deep convolutional neural networks (CNN). The applicability of the proposed mathematical model for processing video information affected by heterogeneous noise and subsequent face recognition has been substantiated.

2. A multi-stage model based on convolutional neural networks has been developed, aimed at improving recognition accuracy and optimizing frame processing time in video streams under conditions of complex illumination, noise presence, and scale variations.

Theoretical significance of the research lies in the development of mathematical models for image recognition in video streams and image processing methods based on neural networks.

Practical significance of the research is determined by the possibility of improving the efficiency of image recognition in real-time systems through the application of the proposed algorithms in video surveillance systems and intelligent video analytics.

Implementation of research results. The results of the dissertation research have been implemented in practice at LLP “QUARES” with the aim of improving the

accuracy and speed of face identification in intelligent video analytics systems under complex lighting conditions.

Approbation of the dissertation results. The main results of the dissertation research were presented and discussed at scientific seminars of the EP “Information Systems” of the Saken Seifullin Kazakh Agrotechnical Research University, Astana, 2026.

International journals with a non-zero impact factor included in the Scopus database – 2 articles. The author’s contribution includes the formulation of the scientific hypothesis, data collection and preprocessing, development and technical implementation of the experimental part of the study, interpretation of the obtained results, and preparation of publications.

1. Mutalova, Z.; Shaushenova, A.; Nurpeisova, A.; et.al. Development of a mathematical model for detecting moving objects in video streams in real-time. IEEE Access. 2024, 10, 136. https://doi.org/10.1109/ACCESS.2024.3487783.
2. Nurpeisova, A., Shaushenova A., Kuznetsov O., Ispussinov A., Mutalova Z., Kassymova A. Deep Residual Learning for Face Anti-Spoofing: A Mathematical Framework for Optimized Skip Connections. Technologies, 2025, Volume 13, Issue 9, 413; https://doi.org/10.3390/technologies13090413.

Articles in journals recommended by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan – 2 articles. In these publications, the author developed the research concept and data analysis methodology, participated in data collection, interpretation of results, and preparation of scientific papers for publication.

1. Mutalova Zh.S., Shaushenova A.G., Issakova G.O., Nurpeisova A.A., Ongarbayeva M.B., Abdykalikova G.A. A method for recognizing a person from a face image based on moving a point along the guides. Proceedings of the National Academy of Sciences of the Republic of Kazakhstan. Physics and Mathematics Series, ISSN 1991-346X Volume 3. Number 351 (2024). 118-137 https://doi.org/10.32014/2024.2518-1726.296.
2. Mutalova Zh.S., Shaushenova A.G., Nurpeisova A.A., Ainagulova A.S., Ongarbayeva M.B. Comparative analysis of occlusive facial recognition methods based on neural networks. University works. – 2026. №1 (102) 374-380 https://doi.org/10.52209/1609-1825_2026_1_374.

In journals included in the Russian Science Citation Index (RSCI) national bibliographic database – 2 articles.

1. Mutalova Zh.S., Shaushenova A.G., Nurpeisova A.A., Ongarbayeva M.B. Research of the YOLO-based facial recognition algorithm // Current issues of modern economics. — 2025. — № 6. [Electronic resource]. URL: https://avse-journal.ru/Editions.
2. Mutalova Zh.S., Shaushenova A.G., Nurpeisova A.A., Ongarbayeva M.B. Convolutional neural network based on face recognition function // Current issues of the modern economy. — 2025. — № 6. [Electronic resource]. URL: https://avse-journal.ru/Editions.

At international conferences – 1 article.

1. Mutalova Zh.S., Shaushenova A. G., Nurpeisova A.A. Convolutional neural network algorithm based on improving deep learning functions // Materials of the International Scientific and Practical Conference "Integration of artificial intelligence into the educational process in the era of global digitalization: strategies,

innovations and cybersecurity issues", dedicated to the 70th anniversary of Academician of the National Academy of Engineering of the Republic of Kazakhstan, Doctor of Technical Sciences, Professor Akhmetov Bakytzhan Srazhatdinovich. — Almaty, April 5, 2024, pp. 1148-1153.

Utility model patent – 1.

1. Mutalova Zh.S., Nurpeisova A.A., Shaushenova A. G., Ispussinov A.M., Zhumaliyeva L. D. A neural network-based face recognition system. The patent for a utility model of the Republic of Kazakhstan № 9870, 2024. [Electronic resource]. URL: https://gosreestr.kazpatent.kz/Utilitymodel/Details?docNumber=409046.

Personal contribution of the author. The study includes the formulation of the image recognition problem in video streams, the development of a concept for applying neural networks to build mathematical models, and an analysis of scientific literature in the fields of computer vision, deep learning, and image processing. The author developed the architecture of an intelligent video data processing system, proposed mathematical models for image processing aimed at reducing the influence of noise and distortions, and substantiated a multi-stage approach combining Retinex algorithms and convolutional neural networks. In addition, deep neural network models were developed to improve image recognition accuracy, mathematical methods for enhancing face recognition quality, and algorithms for increasing the efficiency of video frame processing. As part of the research, a prototype software implementation of the system was developed, data preparation and processing were carried out, computational experiments were conducted, and the obtained results were analyzed. All major results of the dissertation research were obtained personally by the author.

Publications related to the dissertation research. A total of 8 (eight) scientific works were published on the topic of the dissertation research, including: 2 (two) articles in journals recommended by the Committee for Quality Assurance in Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan; 2 (two) articles in international scientific journals indexed in the Scopus database; 2 (two) articles in journals included in the Russian Science Citation Index (RSCI) national bibliographic database; 1 (one) article in international conference proceedings; and 1 (one) utility model patent.

Structure and volume of the dissertation. The dissertation consists of an introduction, five chapters, a conclusion, a list of references, and appendices. The total volume of the dissertation is 148 pages (including 56 figures and 7 tables).

In the **introduction**, the relevance of the research topic is substantiated, related to the need to develop and improve methods of image recognition in video streams based on neural networks, aimed at increasing the accuracy and efficiency of video data processing under changing external factors and noisy environments.

In the first chapter, the main methods of image recognition in video streams were examined, including traditional approaches and deep learning methods. It was established that traditional methods possess high computational efficiency; however, their ability to adapt to complex data conditions is limited. In real-time processing tasks, neural networks demonstrated higher accuracy and robustness. It was also noted that transformer-based models have significant potential, although they require substantial computational resources. An analysis of the main types of neural network architectures used for image processing, as well as their operating principles, was conducted. It was determined that convolutional neural networks (CNNs) are among

the most effective models for solving image recognition, image classification, and object detection tasks.

In the second chapter, a mathematical model for image preprocessing and face recognition based on convolutional neural networks and Retinex algorithms was developed. Within the proposed model, the processes of improving input data quality through non-uniform illumination correction, contrast normalization, and noise suppression were formalized. It was proven that the application of the Retinex algorithm at the preprocessing stage increases the informativeness of the feature space and significantly improves face recognition accuracy, especially in complex visual environments with varying illumination conditions. A multi-stage model aimed at improving image recognition accuracy in video streams and reducing processing time was also developed. The proposed approach includes a comprehensive pipeline consisting of preprocessing, face region detection, feature extraction, and identification stages. The integration of the Retinex algorithm with Faster R-CNN, EfficientNetB0, and Self-Attention mechanisms improved image quality and recognition accuracy. Embedding generation was performed using the FaceNet model.

In the third chapter, a multi-level convolutional neural network architecture aimed at improving data classification accuracy and processing efficiency was developed. In the proposed Nexiface architecture, the Retinex preprocessing module, Faster R-CNN detector, EfficientNetB0 model, Self-Attention feature extraction module, and FaceNet identification model were integrated into a unified processing pipeline. The Retinex algorithm reduced the influence of non-uniform illumination and improved image quality, thereby increasing the stability of the neural network. The two-stage Faster R-CNN detection architecture enabled highly accurate localization of facial regions. Owing to the lightweight EfficientNetB0 architecture, computational costs were reduced while maintaining high processing speed and efficient extraction of significant facial features. The Self-Attention mechanism enabled the model to focus on the most informative regions of the image, such as the eyes, nose, and facial contours, which enhanced the informativeness of the feature space. As a result, a high-performance and robust Nexiface architecture was developed, and its effectiveness was confirmed through practical implementation.

The **conclusion** summarizes the main research results and presents findings confirming the achievement of the research goal and the completion of dissertation tasks. The appendices include a utility model patent and an implementation act.

The dissertation was carried out within the framework of the scientific project AR23486538 “Research and Development of an Artificial Intelligence-Based Object Recognition System in Video Streams”, funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan for 2024–2026, implemented in accordance with the priority directions of scientific and technological development of the Republic of Kazakhstan.