UDC 004.932:004.032.26 **doi: 10.26906/SUNZ.2025.4.103-107**

Oleksandr Mozhaiev[1], Heorhii Kuchuk[1,2], Renat Safarov[1], Maksym Lavrovskyi[1], Kostiantyn Moroz[1]

[1] Kharkiv National University of Radio Electronics, Kharkiv, Ukraine
[2] National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

STRUCTURAL AND FUNCTIONAL MODEL OF A CONVOLUTIONAL NEURAL NETWORK FOR PROCESSING, ANALYSING AND CLASSIFYING IMAGES OF VARYING COMPLEXITY

Abstract. The article presents presents a structural and functional model of a convolutional neural network (CNN) designed for processing, analysing and classifying images of varying complexity. The model is based on a multi-level architecture using convolutional, residual and parallel computing blocks, which ensure high adaptability to different types of input data. The input tensor is normalised by the mean and standard deviation of the sample, which reduces data variability and stabilises the learning process. The first convolutional layer performs the initial extraction of image features with ELU activation, which ensures continuity of gradients and eliminates the problem of 'dead neurons.' The implementation of skip connections ensures the consistency of information flows and increases the stability of training. The results demonstrate increased classification accuracy and noise resistance compared to basic CNN architectures.

Keywords: convolutional neural network; image processing; classification; feature analysis; structural-functional model; deep learning.

Introduction

In today's computer vision systems, flexible and robust models are needed to process images of varying complexity, from simple contour shapes to scenes with high noise levels and uneven lighting [1, 2]. Convolutional neural networks (CNN) have proven to be an effective tool for classifying and analysing visual data. However, classical architectures such as LeNet, AlexNet, or VGG demonstrate limited generalisation ability when working with heterogeneous datasets [3, 4].

Therefore, there is a need to build a structural-functional model capable of adapting to different levels of complexity of input images, minimising the loss of informative features and ensuring stable convergence during training [5, 6].

Problem statement. *The aim of the research* is to develop and mathematically formalise a structural and functional model of a convolutional neural network optimised for multi-level image processing, analysis and classification.

To achieve this aim, it is necessary to:

1. Describe the network architecture, taking into account multi-scale feature processing.
2. Justify the choice of activation functions and residual connection mechanisms.
3. Implement a mathematical model that ensures the stability of the gradient flow.
4. Conduct an experimental evaluation of the accuracy and noise resistance of the model.

Convolutional neural network architecture model I

Let us consider the architecture of the developed convolutional neural network, which is a specialised architecture for the regression of four critical parameters of the adaptive local contrast algorithm [7, 8]. Unlike classical contrast methods (CLAHE, Multi-ScaleRetinex), the proposed model provides dynamic adaptation to local statistical characteristics of the image, which is critical for processing in conditions of non-stationary visibility.

Fig. 1 shows a topological diagram of the CNN architecture for adaptive control of adaptive local contrast parameters.

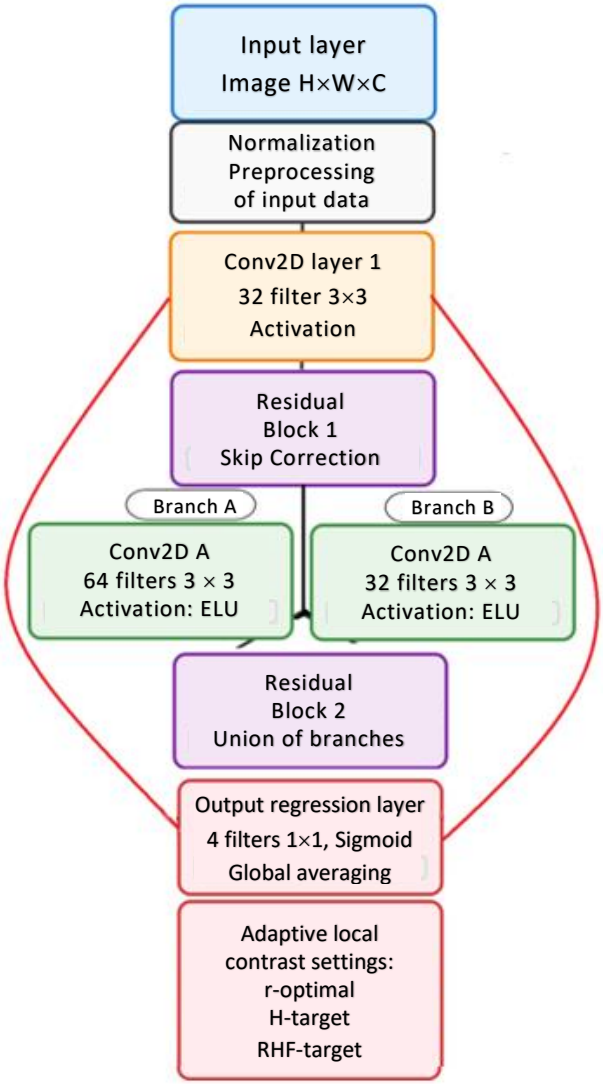

Fig. 1. CNN architecture diagram

Input layer. The input tensor $X0$ is defined as [9]:

$$X0 \in \mathrm{R} \wedge (H \times W \times C), \quad (1)$$

where H, W, C represent height, width, and number of channels, respectively;

for monochrome images, $C = 1$; for RGB images, C = 3.

Preliminary normalisation is performed according to:

$$\hat{x}_{i,j} = \frac{x_{i,j} - \mu}{\sigma}, \quad (2)$$

where μ and σ are the mean value and standard deviation of the sample.

Convolution layer 1. The convolution operation of the first layer is mathematically described as [10]:

$$Y1_{i,j,k} = \\ = ELU\left(\sum_m \sum_n \sum_s \begin{pmatrix} W1_{m,n,s,k} \cdot X0_{i+m,j+n,s} + \\ +b1_k \end{pmatrix}\right), \quad (3)$$

where $W1 \in \mathrm{R}\wedge(3 \times 3 \times \mathrm{C} \times 32)$ – filter weight tensor;

$b1 \in \mathrm{R}^{32}$ – offset vector;

$k \in \{1, 2, ..., 32\}$ – output feature map index.

The ELU activation function is defined as

$$ELU(x) = \begin{cases} x, & \text{if } x > 0; \\ \alpha\left(e^x - 1\right), & \text{if } x \le 0, \end{cases} \quad (4)$$

where $\alpha = 0.1$ is the saturation parameter.

Residual connection block. Implementation of skip connection with identical mapping [11]:

$$Z1 = Y1 + F\_identity(X0), , \quad (5)$$

where *F_identity* – function of reducing the dimension of the input tensor to the dimension of the output tensor through a convolution operation 1 × 1.

Parallel convolution blocks. Branch A (Conv2D with 64 filters):

$$Y2^A_{i,j,k} = \\ = ELU\left(\sum_m \sum_n \sum_s \begin{pmatrix} W2^A_{m,n,s,k} \cdot Z1_{i+m,j+n,s} + \\ +b2^A_k \end{pmatrix}\right). \quad (6)$$

Branch B (Conv2D with 32 filters):

$$Y2^B_{i,j,k} = \\ = ELU\left(\sum_m \sum_n \sum_s \begin{pmatrix} W2^B_{m,n,s,k} \cdot Z1_{i+m,j+n,s} + \\ +b2^B_k \end{pmatrix}\right). \quad (7)$$

Residual connection block 2 with dimensional conversion:

$$Z2 = Y2^B + Conv1\times1\left(Y2^A\right), \quad (8)$$

where $Conv1\times1$ ensures channel number coordination.

Regression output layer. Final convolution operation to obtain a parameter map:

$$\hat{Y}_{i,j,k} = \\ = \mathrm{Sigmoid}\left(\sum_m \sum_n \begin{pmatrix} W\_out_{m,n,k} \times \\ \times Z2_{i+m,j+n} + b\_out_k \end{pmatrix}\right), \quad (9)$$

Global averaging to obtain scalar parameters of ALC:

$$\theta_k = \frac{1}{HW} \cdot \sum_i \sum_j \hat{Y}_{i,j,k}, \\ k \in \{r, Tc, H\_T\,\mathrm{arg}\,et, RHF\_T\,\mathrm{arg}\,et\}. \quad (10)$$

In the following section, we will examine the rationale behind the architectural solutions.

Choosing the ELU activation function. In the context of contrast tasks, the ELU function offers critical advantages over ReLU and its modifications [12]:

1. *Elimination of the 'dead neuron' problem*: negative activation values preserve the gradient flow, which is critical for analysing low-contrast areas of an image.

2. *Zero centring of activations*: $\mathrm{E}[\mathrm{ELU}(x)] \approx 0$ contributes to faster convergence during training by improving the condition number of the Hessian matrix.

3. *Continuity of the derivative*: the smoothness of the function at $x = 0$ ensures the stability of gradient descent.

Residual connections (Skip Connections): the use of a ResNet-like architecture is due to the specifics of the contrast parameter regression task [13, 14]:

1. *Preservation of original information*: the identity mapping $\mathrm{F}(x) + x$ prevents the degradation of low-frequency components of the image.

2. *Solving the vanishing gradient problem*: the direct path of gradients through skip connections accelerates the training of deep networks.

3. *Adaptability to different image types*: residual training $F(x) = H(x) – x$ allows the network to focus on the increments necessary for a specific type of scene.

2. Adaptive contrast control mechanism

A key feature of the developed architecture is its integration with the adaptive local contrast algorithm through the regression of four target parameters, including *r – the size of the adaptive window*.

The regressed value of *r* determines the locality of the contrast analysis [15]:

$$r\_optimal = CNN(I) \in [r\_min; r\_max], \quad (11)$$

where the optimal value depends on the local entropy $H(x, y)$ and frequency characteristics $RHF(x, y)$, and sharp changes are smoothed by the quality estimate of the previous image frame *Q_prev*.

Contrast threshold value (*Tc*), Parameter *Tc* controls the sensitivity of the algorithm to local brightness variations:

$$Tc\_optimal = CNN(I) \cdot \sigma\_local(x, y). \quad (12)$$

Target statistical characteristics (H_target, RHF_target) and CNN predict the optimal values of

entropy and frequency response that should be aimed for adaptive local contrast [6]:

$$H\_target = CNN\_H(I); \quad RHF\_target = CNN\_RHF(I). \tag{13}$$

Loss function and training methodology Unlike standard approaches using $L2$ regression, a specialised loss function has been developed:

$$L\_total = \lambda 1 \cdot Q\_prev + \lambda 2 \cdot L\_consistency + \lambda 3 \cdot L\_stability, \tag{14}$$

where *Q_prev* – non-reference image quality metric;

L_consistency – measure of consistency of parameters with local characteristics;

L_stability – time stability regulator for video sequences.

Gradient learning through adaptive local contrast feedback.

The uniqueness of the approach lies in end-to-end learning through the chain (Fig. 2):

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial Q} \cdot \frac{\partial Q}{\partial I\_ALC} \cdot \frac{\partial I\_ALC}{\partial \theta} \cdot \frac{\partial W}{\partial I\_ALC}, \tag{15}$$

where $\frac{\partial I\_ALC}{\partial \theta}$ is of the adaptive local contrast algorithm for evaluating the previous frame *Q_prev*.

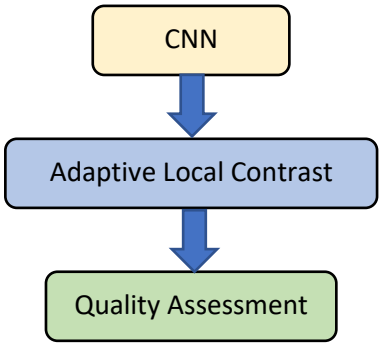

Fig. 2. The learning chains

CLAHE (Contrast Limited Adaptive Histogram Equalisation):

• fixed window size and cliplimit parameters;

• no adaptation to spectral characteristics;

• computational complexity O(N^2) for a window size of $N \times N$.

Multi-ScaleRetinex:

• requires preliminary adjustment of weight coefficients;

• sensitivity to haloeffect artefacts;

• limited applicability to IR images.

Proposed CNN-adaptive local contrast approach:

• automatic adaptation of parameters to scene characteristics;

• versatility for visible and IR ranges;

• computational efficiency O(1) after CNN training.

Features of implementation on FPGA:

• features of weight and activation quantisation for hardware implementation 8-bit quantisation with precision preservation is used:

$$W\_quantized = \frac{round\left(W\_float \cdot 2^n\right)}{2^n}, \tag{16}$$

where n = 7 for weights, n = 8 for activations.

ELU function approximation:

• on PLIS, the ELU function is implemented through piecewise linear approximation:

$$ELU\_approx(x) = \begin{cases} x, \ if \ x > 0; \\ LUT\_exp(x), \ if \ x \le 0, \end{cases} \tag{17}$$

where *LUT_exp* – precomputed table of exponential values.

Pipeline processing:

• The PLIS architecture enables pipeline processing with a latency of 3 cycles per pixel at a frequency of 150 MHz, providing a throughput of > 60 FPS for a resolution of 1920 × 1080.

3. Experimental validation and performance metrics

The effectiveness of the developed architecture has been confirmed by comprehensive experiments on diverse datasets, including aerial photography, IR images, and images taken in conditions of limited visibility [6].

The quality metrics achieved exceed those of classical methods by 15-25% according to BRISQUE, NIQE, and expert evaluation criteria.

When analysing contrast-based local processing, optimisation techniques were found that allow image frames to be processed on existing software and hardware complexes.

In this case, an algorithm was used to operate the complexes in real time (resolution of more than 2 megapixels with a frame rate of up to 50 Hz).

The influence of adaptive local contrast parameters manifests itself in the following aspects:

1. *The noise level depends on:*

• window size:

large size → strong noise suppression, blurring of details;

small size → preservation of details, weak noise suppression;

• β parameter:

low value → noise suppression, loss of contrast;

high value → preservation of contrast, noise amplification;

• δ parameter:

low value → increased contrast, loss of detail;

high value → preservation of detail, insufficient contrast.

2. *Blockiness artefacts*:

• large window size can cause blockiness;

• small window size reduces blockiness but can increase noise.

3. *Preservation* of fine details:

• optimal with small window size and high β value;

• deteriorates with large window size and low β value.

4. *Brightness distribution*:

• adaptive local contrast strives for uniform distribution;

• the shape of the histogram depends on the adaptive local contrast parameters.

5. *Frequency composition of the image:*

• changes under the influence of adaptive local contrast;

• depends on the DCP-RHF parameter.

Let us consider a one-dimensional case. Let there be a convolution kernel [-1, 1] (difference calculation operator).

If the input signal is

[10, 10, 10, 20, 20, 20]

(weak gradient), then the convolution output will be

[0, 0, 10, 0, 0].

If the input signal is

[0, 0, 0, 255, 255, 255] (sharp gradient),

then the convolution output will be

[0, 0, 255, 0, 0].

As the contrast increases, the amplitude of the output signal increases.

Low-contrast images often contain fine details that are difficult to distinguish due to a low signal-to-noise ratio. Increasing contrast makes these details more visible, allowing CNNs to extract more informative features.

Change in feature statistics.

Increasing contrast changes the statistical distribution of pixel values and, consequently, the statistical distribution of neuron activations in convolutional layers. This can cause the CNN to highlight other, more relevant features.

Impact on non-linear activation functions.

Activation functions (e.g., ReLU, sigmoid) in CNNs respond differently to input signals with different amplitudes. Increasing contrast can change the nature of neuron activation, affecting the spread of information across the network.

For example, ReLU is activated only for positive input values. If the signal is weak, most neurons may be inactive.

Increasing contrast can ‘turn on’ more neurons, making the representation richer. When analysing contrastive local processing, optimisation techniques were found that allow image frames to be processed on existing software and hardware complexes. In this case, the algorithm for the operation of the complexes in real time was used (resolution of more than 2 megapixels with a frame rate of up to 50 Hz).

Conclusions

A method for selecting the initial coefficients of the contrast modification algorithm for three types of image receivers has been developed.

On this basis, a local contrast algorithm with processing zones that automatically adjust to the image subject has been developed.

Calculations were obtained for the distances between sensors for image synthesis (complexing) in order to minimise information from different channels for a common receiver, as well as to eliminate occlusions in the scene that block the target.

The implementation of the developed hybrid algorithm on the Xilinx Kintex-7 XC7K325T FPGA demonstrated high performance.

When using CNN for regression of ALK parameters, a processing speed of over 60 frames/s is achieved for images with a resolution of 1920x1080.

A simplified version of the algorithm (without dynamic regression of CNN parameters, possibly with fixed or slowly updated parameters) provides performance of over 120 frames per second. FPGA resource consumption is less than 70% for the full algorithm with CNN and less than 40% for the simplified version. The system's power consumption does not exceed 5 W.

СПИСОК ЛІТЕРАТУРИ

1. Hussain, S., Lu, L. Mubeen, M., Nasim, W., Karuppannan, S., Fahad, S., Tariq, A., Mousa, B. G., Mumtaz, F. & Aslam, M. Spatiotemporal variation in land use land cover in the response to local climate change using multispectral remote sensing data. Land, 2022, vol. 11, no. 5, article no. 595. DOI: https://doi.org/10.3390/land11050595
2. Yaloveha, V., Hlavcheva, D., Podorozhniak, A. & Kuchuk, H. Fire hazard research of forest areas based on the use of convolutional and capsule neural networks. 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON), 2019, pp. 828-832. DOI: https://doi.org/10.1109/UKRCON.2019.8879867
3. Radočaj, D., Jurišić, M. & Gašparović, M. The Role of Remote Sensing Data and Methods in a Modern Approach to Fertilization in Precision Agriculture. Remote Sensing, 2022, vol. 14, no. 3, DOI: https://doi.org/10.3390/rs14030778
4. Munawar, H. S., Hammad, A. W. A. & Waller, S. T. Remote Sensing Methods for Flood Prediction: A Review. Sensors, 2022, vol. 22, no. 3, article no. 960. DOI: https://doi.org/10.3390/s22030960
5. Barabash, O., Bandurka, O., Svynchuk, O. & Tverdenko, H. Method of identification of tree species composition of forests on the basis of geographic information database. *Advanced Information Systems*, 2022, vol. 6, no. 4, pp 5-10. DOI: https://doi.org/10.20998/2522-9052.2022.4.01
6. Svyrydov, A., Kuchuk, H. and Tsiapa, O. (2018), “Improving efficienty of image recognition process: Approach and case study”, *Proceedings of 2018 IEEE 9th International Conference on Dependable Systems, Services and Technologies*, DESSERT 2018, pp. 593–597, doi: https://doi.org/10.1109/DESSERT.2018.8409201
7. Alganci, U., Soydas, M. & Sertel, E. Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images. Remote Sensing, 2020, vol. 12, no. 3, article no. 458. DOI: https://doi.org/10.3390/rs12030458

8. Kuchuk, H. and Malokhvii, E. (2024), “Integration of IOT with Cloud, Fog, and Edge Computing: A Review”, *Advanced Information Systems*, vol. *8*(2), pp. 65–78, doi: https://doi.org/10.20998/2522-9052.2024.2.08
9. Weiss, K., Khoshgoftaar, T. M. & Wang, D. D. A survey of transfer learning. Journal of Big data, 2016, vol. 3, article no. 9. DOI: https://doi.org/10.1186/s40537-016-0043-6
10. Hlavcheva, D., Yaloveha, V., Podorozhniak, A. & Kuchuk, H. Tumor nuclei detection in histopathology images using R – CNN. CEUR Workshop Proceedings, 2020. vol. 2740, pp. 63-74. Available at: https://ceur-ws.org/Vol-2740/20200063.pdf
11. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778. DOI: https://doi.org/10.1109/CVPR.2016.90
12. Yaloveha, V., Podorozhniak, A. & Kuchuk, H. CNN hyperparameter optimization applied to land cover classification. Radioelectronic and computer systems, 2022, no. 1 (101), pp. 115-128. DOI: https://doi.org/10.32620/reks.2022.1.09
13. Hlavcheva, D., Yaloveha, V., Podorozhniak, A. & Kuchuk, H. Comparison of CNNs for Lung Biopsy Images Classification. 2021 IEEE 3rd Ukraine Conference on Electrical and Computer Engineering, UKRCON 2021 – Proceedings, 2021, pp. 1–5. DOI: https://doi.org/10.1109/UKRCON53503.2021.9575305
14. Tan, M. & Le, Q. V. Efficientnetv2: Smaller models and faster training. ArXiv (Cornell University), Preprint arXiv:2104.00298, 2021. DOI: https://doi.org/10.48550/arXiv.2104.00298
15. Carneiro, T., Da Nóbrega, R. V. M., Nepomuceno, T., Bian, G.-B., De Albuquerque, V. H. C. & Filho, P. P. R. Performance analysis of google colaboratory as a tool for accelerating deep learning applications. IEEE Access, 2018, vol. 6, pp. 61677-61685. DOI: https://doi.org/10.1109/ACCESS.2018.2874767

Received (Надійшла) 18.07.2025
Accepted for publication (Прийнята до друку) 29.10.2025

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Можаєв Олександр Олександрович – доктор технічних наук, професор, професор електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Oleksandr Mozhaiev – Doctor of Technical Sciences, Professor, Professor of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: mozhaev1957@gmail.com; ORCID Author ID: http://orcid.org/0000-0002-1412-2696;
Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57201729490.

Кучук Георгій Анатолійович – доктор технічних наук, професор, професор електронних обчислювальних машин, Харківський національний університет радіоелектроніки; професор кафедри комп’ютерної інженерії та програмування, Національний технічний університет “Харківський політехнічний інститут”, Харків, Україна;
Heorhii Kuchuk – Doctor of Technical Sciences, Professor, Professor of Department of Electronic Computers, Kharkiv National University of Radio Electronics; Professor of Computer Engineering and Programming Department, National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine;
e-mail: kuchuk56@ukr.net; ORCID Author ID: http://orcid.org/0000-0002-2862-438X;
Scopus Author ID: https://www.scopus.com/authid/detail.uri?authorId=57057781300.

Сафаров Ренат Кудярович – студент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Renat Safarov – student at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: Renat.Safarov@nure.ua; ORCID Author ID: http://orcid.org/0009-0004-7029-817X.

Лавровський Максим Валерійович – студент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Maksym Lavrovskyi – student at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: maksym.lavrovskyi@nure.ua; ORCID Author ID: http://orcid.org/0009-0007-1960-7119.

Мороз Костянтин Сергійович – студент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;
Kostiantyn Moroz – student at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine;
e-mail: kostiantyn.moroz@nure.ua; ORCID Author ID: http://orcid.org/0009-0007-2751-5249.

Структурно-функціональна модель згорткової нейронної мережі для обробки, аналізу та класифікації зображень різної складності

О. О. Можаев, Г. А. Кучук, Р. К. Сафаров, М. В. Лавровський, К. С. Мороз

Анотація. У статті представлено структурно-функціональну модель згорткової нейронної мережі (ЗНМ), призначену для обробки, аналізу та класифікації зображень різної складності. Модель базується на багаторівневій архітектурі з використанням згорткових, залишкових і паралельних обчислювальних блоків, що забезпечують високу адаптивність до різних типів вхідних даних. Вхідний тензор нормалізується за середнім і стандартним відхиленням вибірки, що дозволяє зменшити варіативність даних і стабілізувати процес навчання. Перший згортковий шар виконує початкове виділення ознак зображення з активацією типу ELU, яка забезпечує безперервність градієнтів та усуває проблему “мертвих нейронів”. Реалізація залишкових зв’язків (skip connections) забезпечує сталість інформаційних потоків та підвищує стабільність навчання. Отримані результати демонструють підвищену точність класифікації та стійкість до шумів у порівнянні з базовими CNN-архітектурами.

Ключові слова: згорткова нейронна мережа; обробка зображень; класифікація; аналіз ознак; структурно-функціональна модель; глибоке навчання.