

UDC 004.032.26

Kozlov V.¹, Subbotin S.², Oliinyk A.², Leoshchenko S.³

¹PhD student NU «Zaporizhzhia Polytechnic»

²Professor NU «Zaporizhzhia Polytechnic»

³Associate professor NU «Zaporizhzhia Polytechnic»

A SURVEY OF MODERN APPROACHES TO IMAGE CLASSIFICATION AND ANALYSIS OF THE TRANSITION FROM CONVOLUTIONAL TO HYBRID ARCHITECTURES

The rapid advancement of computer vision has led to a growing demand for accurate and efficient image classification methods. This has driven continuous research into model architectures and learning paradigms, reflecting broader trends in data representation, computational capabilities, and performance requirements.

The development of image classification methods has followed a staged evolution. Early approaches were dominated by classical computer vision

algorithms based on handcrafted feature extraction, such as SIFT, SURF, and HOG, typically combined with machine learning methods like SVM. These approaches achieved satisfactory performance on relatively simple tasks. However, they showed limited generalization ability and struggled to scale to complex scenes and large datasets.

The next stage involved the application of general-purpose artificial neural networks, particularly multilayer perceptrons, followed by early convolutional neural networks such as LeNet. These models introduced the concept of end-to-end learning, enabling automatic feature extraction without manual feature descriptor design. However, their practical use was limited by shallow network depth and the lack of sufficient computational resources at the time.

A major breakthrough in image classification came with the emergence of deep convolutional neural networks. The AlexNet architecture marked the beginning of a new era characterized by increased model depth and complexity. Subsequent architectures, including VGG, GoogLeNet, and ResNet, focused on improving classification accuracy through deeper structures, advanced architectural components, and more stable training processes. This stage led to a significant reduction in error rates on benchmark datasets such as ImageNet, but it also increased computational demands and the number of model parameters.

Subsequent research shifted toward optimizing architectures to reduce memory usage and computational cost while maintaining high accuracy. This led to the development of efficient models such as SqueezeNet, MobileNet, EfficientNet, and RegNet. These approaches applied techniques such as factorized convolutions, bottleneck blocks, systematic scaling, and neural architecture search. They achieved a balance between accuracy and efficiency, which is critical for real-world applications, especially in embedded and mobile systems [1].

The next stage involved the adaptation of transformer-based architectures for computer vision tasks. The Vision Transformer demonstrated that self-attention mechanisms can effectively model global spatial relationships, while the Swin Transformer introduced hierarchical processing to reduce computational cost. Despite their high accuracy, transformer-based models often require large training datasets and have weaker built-in assumptions about visual structure compared to convolutional networks.

The current stage is characterized by a shift toward hybrid and multi-architecture approaches that combine the strengths of convolutional and transformer-based models. Architectures such as CoAtNet, ConvNeXt, and EVCC demonstrate the integration of local feature extraction from convolutional networks, global context modeling from transformers, and efficient scaling strategies [2]. This approach enables high classification accuracy with moderate computational cost.

In conclusion, the evolution of image classification methods shows a clear transition from separate architectural paradigms to integrated solutions. Future progress in computer vision is likely to depend not on competition between different model types, but on their effective combination, adaptability to specific tasks, and efficient use of computational resources.

REFERENCES

1. Szeliski R. Computer Vision: Algorithms and Applications, 2nd ed. URL : <https://szeliski.org/Book> (Accessed: 21.03.2026).
2. Biswas, A. The History of Convolutional Neural Networks for Image Classification (1989 - Today). URL : <https://medium.com/data-science/the-history-of-convolutional-neural-networks-for-image-classification-1989-today-5ea8a5c5fe20> (Accessed: 21.03.2026).