ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

INFORMATION TECHNOLOGY

DOI: 10.20998/2079-0023.2026.01.07
UDC 004.932:004.8

S. V. KOVALENKO, Candidate of Technical Sciences, Docent, National Technical University "Kharkiv Polytechnic Institute", Professor of the Department of System Analysis and Information-Analytical Technologies; Kharkiv, Ukraine; e-mail: serhii.kovalenko@khpi.edu.ua; ORCID: https://orcid.org/0000-0001-8763-0862
S. M. KOVALENKO, Candidate of Technical Sciences, Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor of the Department of Software Engineering and Management Intelligent Technologies; Kharkiv, Ukraine; e-mail: svitlana.kovalenko@khpi.edu.ua; ORCID: https://orcid.org/0000-0001-6770-6778
T. E. ALEKSANDROVA, Doctor of Technical Sciences, Professor, National Technical University "Kharkiv Polytechnic Institute", Head of the Department of System Analysis and Information-Analytical Technologies; Kharkiv, Ukraine; e-mail: tetyana.aleksandrova @khpi.edu.ua; ORCID: https://orcid.org/ 0000-0001-9596-0669
M. M. MALKO, Candidate of Technical Sciences, Docent, National Technical University "Kharkiv Polytechnic Institute", Professor of the Department of System Analysis and Information-Analytical Technologies; Kharkiv, Ukraine; e-mail: maxim.malko@khpi.edu.ua; ORCID: https://orcid.org/ 0000-0002-0125-2141

AUTOMATED OBJECT SEGMENTATION AND BACKGROUND REMOVAL IN MULTIMEDIA CONTENT PROCESSING SYSTEMS

In today's rapidly developing digital technologies, image processing occupies a key place in the fields of computer vision and graphics. One of the most popular tasks is the automatic separation of foreground objects from the background, which is of critical importance for web design, e-commerce, the advertising industry, and augmented reality systems. Traditional manual editing methods are laborious and dependent on user experience, especially when working with complex textures or fuzzy object boundaries. The development of deep learning methods make it possible to automate these processes, providing high accuracy and speed of processing in real time. The aim of the work is to analyze existing algorithms for automatic background removal and develop an effective methodology for object segmentation using modern computer vision models to improve the quality of multimedia content. The work uses a comprehensive approach based on the use of the hybrid neural network architecture BEN2. The key feature of the method is the implementation of an innovative Confidence Guided Matting pipeline, which implements a two-stage Bayesian approach: first, coarse segmentation (BEN Base) is performed to create a "draft" mask, and then targeted refinement of boundaries in areas of low confidence of the model (BEN Refiner). As part of the research, a software application based on the Streamlit framework was developed, which provides automated inference of the BEN2 model. The system supports loading images in popular graphic formats, batch data processing, and visualization of results using an interactive slider. The high efficiency of the algorithm was experimentally confirmed: for complex objects, the architecture provides accurate matting of hair and small details, minimizing artifacts such as "ragged edges". The processing workflow was optimized by using local caching of models and supporting acceleration on GPU/CPU. The proposed approach based on the BEN2 architecture and boundary refinement mechanisms demonstrated higher accuracy compared to classical one-stage segmentation methods. The developed system is scalable and suitable for integration into real multimedia processing information systems, which allows significantly reducing the time for preparing graphic content and increasing its professional level.
Keywords: multimedia content processing, automatic background removal, image segmentation, deep learning, alpha matting, BEN2 architecture.

Introduction. In today's rapidly developing digital technologies and multimedia systems, image processing occupies an important place in the fields of computer vision, graphics, and information technology. One of the key tasks of visual data processing is the automatic separation of foreground objects from the background [1], which has wide practical applications in web design, e-commerce, digital photography, the advertising industry, and augmented reality systems.

Traditional background removal methods involve manual image editing, which is laborious, requires specialized skills, and is time-consuming. Processing images with complex background textures, uneven lighting, or blurred object boundaries is particularly challenging. In such cases, the accuracy and quality of the result largely depend on the user's experience.

With the development of machine learning methods and deep neural networks, effective approaches to automating the background removal process have emerged. Modern algorithms enable high-precision image segmentation, ensuring rapid object detection even in complex scenes.

The relevance of the research is due to the need to improve the efficiency and quality of automatic background removal in images, as well as the development of universal algorithms capable of working in real time and adapting to different types of input data.

Current State of the Problem. The background of an image is an important visual component that significantly affects the user's perception of graphic content [2]. Excessive saturation, complexity, or the presence of extraneous details in the background can lead to distraction and a decrease in the informational value of the main object. This is especially true for applied tasks, such as creating advertising materials, profile images, or content for web resources, where a clear visual hierarchy is required.

Modifying or replacing the background allows you to improve the aesthetic characteristics of the image, ensure focus on the target object and achieve stylistic consistency in a series of graphic materials. In addition, background correction helps to improve the adaptability of images to different display formats (web, printed materials) and ensures compliance with specified thematic or design requirements. Thus, changing the background acts not only as a means of visual editing, but also as a tool for improving the quality and professional level of digital content.

The study of the problem of removing and replacing the background of images is based on a wide range of scientific sources, covering both classical methods of digital image processing and modern approaches to computer vision using deep learning. The fundamental principles of image segmentation are outlined in works devoted to image analysis and pattern recognition. In [3, 4], basic methods of image processing are considered, including threshold segmentation, filtering and contour extraction. These approaches became the basis for the further development of more complex algorithms.

With the development of machine learning methods, considerable attention is paid to the use of convolutional neural networks for segmentation tasks. An important milestone was the introduction of the U-Net architecture [5], which provides high segmentation accuracy due to the symmetric encoder-decoder structure and the use of skip connections. Further research, such as works [6, 7], is aimed at the development of fully convolutional networks (FCNs) and SegNet-type architectures.

Modern sources also highlight the use of hybrid models that combine convolutional networks with attention mechanisms or transformer architectures. In [8], the Vision Transformer (ViT) is presented, which demonstrates high efficiency in image analysis tasks due to the global contextual representation of data.

A separate group of sources is made up of applied research devoted to automatic background removal in real-world conditions [9]. Such works analyze the problems of complex scenes, non-uniform illumination, and translucent objects, and also propose methods for increasing segmentation accuracy by using additional features and post-processing.

Thus, the analysis of scientific sources indicates the evolution of approaches to solving the segmentation problem: from classical image processing algorithms to modern deep learning methods. Despite significant achievements, the issues of increasing the accuracy, speed, and versatility of background removal algorithms remain relevant, which determines the feasibility of further research in this area.

Problem Statement. The relevance of this research is due to the need to develop effective algorithms for automatic background removal that combine high accuracy, speed, and versatility of application in conditions of limited resources and a variety of input data.

The aim of the work is to analyze existing algorithms for automatic background removal in images and develop an effective methodology for object segmentation using modern computer vision and machine learning methods.

To achieve the goal, it is necessary to solve the following tasks:

- to formalize the problem of background selection in digital images;
- conduct research on classical digital processing methods and modern algorithms based on deep learning for background removal;
- to justify the use of the hybrid neural network model BEN2 (Background Erase Network) [10] as a basic tool for high-precision segmentation;
- ensure the implementation of the Confidence Guided Matting (CGM) pipeline for targeted processing of areas where the model has a low level of confidence;
- to justify the use of a combined loss function that combines weighted BCE, IoU, and SSIM metrics to improve local and global segmentation features;
- create a practical application based on the Streamlit framework;
- to evaluate the effectiveness of the developed approach using object accuracy and edge sharpness metrics in various application scenarios.

The object of research is the process of processing digital images in the tasks of segmentation and background removal.

The subject of research is algorithms and methods for automatic background removal in images, in particular approaches based on computer vision and deep learning.

Main Section. Removing the background from an image is widely used in many areas:

- e-commerce and advertising – photos of products on a white or neutral background for online stores, advertising banners and product catalogs;
- graphic design – creating collages and posters, developing logos and branding materials, infographics and presentations;
- photography and media – portrait photos for documents or social networks, replacing the background with a more aesthetic or thematic one, photo processing for magazines and news sites;
- video conferencing and streaming – virtual background in Zoom, Teams, Google Meet, chroma key effects without a green screen;
- gaming industry and 3D applications – cutting sprites and textures for games, preparing references for 3D modeling;
- medicine and science – isolation of objects in microscopic images, processing of medical images (X-ray, MRI);
- documents and identification – passport photos, visas, passes, digital signatures and seals;

- fashion and beauty – photos of clothes and accessories for lookbooks, image processing for fashion websites.

The process of image background removal is based on segmentation tasks, which can be implemented using different methods depending on the complexity of the scene and the chosen software tool. The main approaches include:

1) Manual segmentation – involves the user selecting the object using precise geometric marking tools (e.g., polygonal contours or Bezier curves). This approach provides high accuracy, but is time-consuming and depends on the operator's skills.

2) Automated segmentation – implemented using computer vision and artificial intelligence methods, in particular convolutional neural networks (CNN), which perform pixel classification and object separation from the background without user intervention. This approach is characterized by high processing speed and ease of use.

3) Combined methods – combine automatic segmentation with subsequent manual refinement of the result, which allows achieving a balance between speed and accuracy.

The choice of method depends on the requirements for the quality of the result: for mass image processing, it is advisable to use automated approaches, while for tasks requiring high accuracy (for example, catalog products or design), manual or combined processing is preferred.

The effectiveness of background removal is largely determined by the characteristics of the input image. To achieve a high-quality result, it is advisable to consider the following recommendations:

- use images with uniform and sufficient lighting that ensures the clarity of the object's contours;
- avoid using images with low resolution or significant blur;
- provide contrast between the object and the background to simplify the segmentation task;
- use graphic formats with alpha channel support to preserve transparency;
- when processing manually, use scaling to increase the accuracy of boundary selection.

The work uses a comprehensive approach to solving the problem of automatic background removal, which includes computer vision, mathematical modeling, and deep learning methods. The basis of the research is image segmentation methods using the hybrid BEN2 architecture.

BEN2 implements a new approach to foreground segmentation through an innovative CGM pipeline. The model does not simply “draw a mask” around the object, but first analyzes where it is uncertain, and only then refines these areas in a targeted manner. The BEN architecture consists of two components: BEN Base is used for initial segmentation (makes a first “draft”: divides the image into foreground and background) and BEN Refiner for confidence-based refinement, while purposefully processing pixels where the base model demonstrates a lower level of confidence, which gives more accurate and reliable matting results.

The Multi-view Aggregation Network (MVANet) shows significant potential in dichotomous image segmentation (DIS) using a multi-view learning approach. The BEN2 architecture takes this architecture as a base, but introduces important changes. BEN Base repeats MVANet, but with notable changes: the activation function is replaced (GELU – Gaussian Error Linear Unit is used as the activation function instead of ReLU and PReLU from the original MVANet) and normalization (instance normalization is chosen) [11].

Let the input image be given:

$$I \in \mathbb{R}^{H \times W \times C}, \tag{1}$$

where H, W are image height and width;

C is number of channels (usually $C = 3$ is True Color format for RGB model).

The task of segmentation is to find a function

$$f_\theta(I) = M, \tag{2}$$

where f_θ is parameterized model;

θ is a set of parameters;

$M \in \{0,1\}^{H \times W}$ is a binary mask that separates the object (1) from the background (0).

Final image with transparent background:

$$F = I \odot M, \tag{3}$$

where $\odot$ is element-wise multiplication.

In segmentation and background removal tasks, the key step is to determine the exact boundaries between the foreground object and the background. For this, the BEN2 architecture uses the alpha matting operation, which allows you to simulate semi-transparent transitions between image areas and improve the quality of contour extraction. In this case, we are dealing with real edges that are not binary – each pixel is described by the equation:

$$I_p = \alpha_p F_p + (1 - \alpha_p) B_p, \quad \alpha_p \in [0,1], \tag{4}$$

where F_p is foreground color;

B_p is background color;

α_p is opacity mask.

From (4) the problem arises – to estimate α_p for each pixel. To solve the problem, the model predicts using a trained neural network $\hat{M} = f_\theta(I)$.

In the BEN2 architecture, the matting operation is integrated as a separate boundary refinement module, which works on the basis of a previously obtained segmentation map. The main stages are:

1) Coarse segmentation – in the first stage, the convolutional neural network forms a confidence map (probabilistic map of the object)

$$C_p = 1 - 2\left|\hat{M}_p - 0.5\right|. \tag{5}$$

The closer $\hat{M}_p$ up to 0.5, the lower the confidence.

2) Border region selection – an uncertainty region (trimap) is determined, which includes: a clear foreground ($\alpha \approx 1$), clear background ($\alpha \approx 0$) and an undefined region (transition zone).

3) Alpha channel estimation – for pixels in an undefined area, a neural network or specialized submodule estimates the value α taking into account local textural features, intensity gradients, and contextual information.

4) Boundary enhancement – the resulting alpha mask is used to smooth contours and eliminate artifacts such as "ragged edges".

To train the matting module in BEN2, a combined loss function is used, combining three components: weighted BCE (Binary Cross Entropy), weighted IoU (Intersection-over-Union), and weighted SSIM (Structural Similarity Index):

$$L = \alpha_1 L_{\text{BCE}} + \alpha_2 L_{\text{IoU}} + \alpha_3 L_{\text{SSIM}} . \quad (6)$$

It is designed to capture local, global, and token-level segmentation features.

Binary cross-entropy (local pixels) is calculated using

$$L_{\text{BCE}} = -\sum_p \omega_p \left(M_p \log M_p + (1 - M_p) \log(1 - \hat{M}_p) \right). \quad (7)$$

Intersection over Union (global pixels) is calculated

$$L_{\text{IoU}} = 1 - \frac{\sum_p \omega_p \hat{M}_p M_p}{\sum_p \omega_p \left(\hat{M}_p + M_p - \hat{M}_p M_p \right)} , \quad (8)$$

where $\omega_p = 1 + \gamma(1 - C_p)$ is weight, larger for uncertain pixels.

BEN2 was trained on DIS5K [12] and its own proprietary dataset of 22,000 images for segmentation. The improved model demonstrates improved performance in hair matting, 4K processing, object segmentation, and edge refinement.

Thus, BEN2 [13] is a two-stage Bayesian approach that includes first a coarse segmentation with a confidence assessment, and then targeted refinement where the model is "in doubt." This is what makes this approach more efficient than classical one-stage methods.

After removing the background, you can perform post-processing of the object boundaries. In particular, it is recommended to apply feathering or edge filtering techniques to eliminate artifacts such as "torn" contours or residual background pixels.

Fig. 1 shows a block diagram of the system, illustrating its step-by-step algorithm of operation.

The process begins with launching the application, followed by loading the interface. After loading the image, the system checks whether the model is already available locally. If the answer is negative, the model is downloaded, if positive, the local copy is used. This stage is critical for optimizing response time and reducing network load.

For each image, the BEN2 model is inferred, after which the result is converted to RGBA format (alpha channel extraction). The obtained results are stored in the internal session state, which allows avoiding repeated calculations and maintaining UI interactivity. The system displays the results using an interactive "before/after" slider.

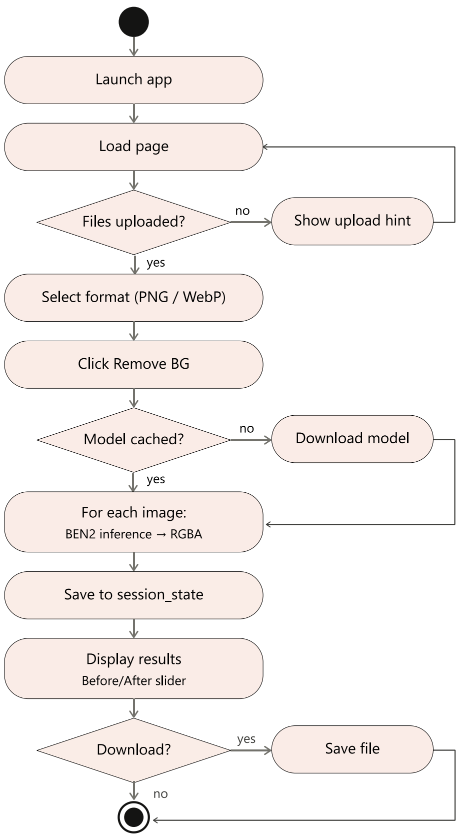

Fig. 1. Flowchart of the system operation based on the BEN2 model architecture

The activity diagram (Fig. 2) details the interaction between three main components: the user, the interface (Streamlit) and the processing module. The user performs the following sequence: opening the application, loading images, selecting the format, starting processing. These actions initiate further processes in the system.

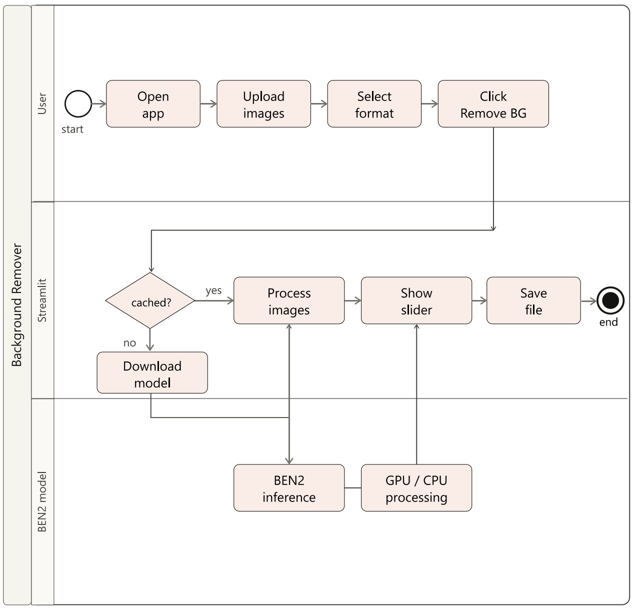

Fig. 2. Activity diagram of the automated background removal process

Streamlit-level logic accepts input data, passes it to the computational module, coordinates inference calls, and displays the results.

Processing at the Background Remover level involves checking the cache, loading the model (if necessary), performing batch image processing (Process images). The BEN2 module receives the image, performs foreground segmentation, uses the GPU or CPU (GPU/CPU processing), returns the result as a mask or RGBA image. In the final step, the results are transmitted to the UI, an interactive slider is displayed, and the user can save the files.

The data flow between the program components is shown in Fig. 3.

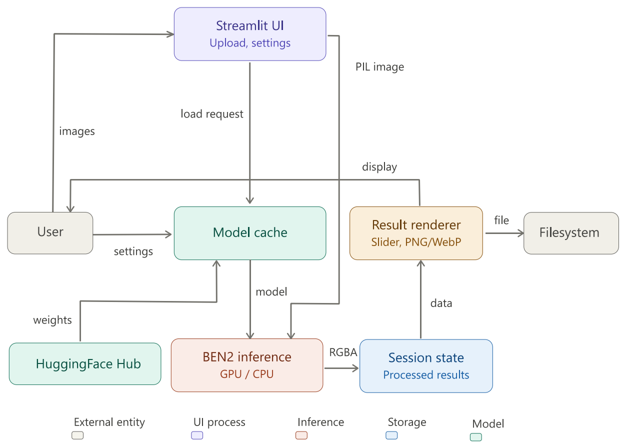

Fig. 3. Data flow between application components

The system architecture is built on a modular principle and includes components of the user interface, data processing, model inference, and results storage.

The user interface is implemented using the Streamlit web framework [14] and provides: loading images; setting processing parameters (output file format, computing device); initiating the processing process (forming a load request); displaying results in the form of an interactive comparison. Loaded images are transferred as PIL (Python Imaging Library) format objects, which provides unified processing of graphic data.

The Model cache module performs the function of centralized management of the model and its weights. The main tasks it solves are: loading model weights from an external repository (HuggingFace Hub); caching the model in RAM for reuse; transferring the model to the inference module; processing parameters received from the user. The use of caching allows you to significantly reduce the time of recalculations and increase system performance.

The central component of the system is the inference module, which implements image processing using the BEN2 model. The main functions: receiving a pre-processed image; performing object segmentation; forming an alpha mask; using hardware acceleration (GPU via CUDA or CPU). The result of the module is an image with a removed background in RGBA format, where the alpha channel is responsible for transparency.

The Session state module is responsible for temporarily storing processing results within the user session. It allows you to accumulate processed images; transfer data to the display module; support batch processing (multiple images); save intermediate results for repeated access. This module allows you to avoid repeated calculations when the user interacts with the interface.

The Result renderer module is responsible for the visualization and export of results. Its capabilities include: displaying results as an interactive slider; preparing files for saving; transferring results to the file system. Interaction with the file system allows the user to save results locally.

Experiment. As part of the research, a software application was developed for automated background removal from digital images based on modern computer vision and deep learning methods. The system is based on the BEN2 neural network model, which provides high-precision segmentation of objects with further refinement of boundaries using alpha matting mechanisms.

The data upload module allows the user to import images via a drag-and-drop interface or select files from the file system. Popular graphic formats (JPG, JPEG, PNG, WEBP, BMP) are supported with a file size limit of 200 MB (Fig. 4).

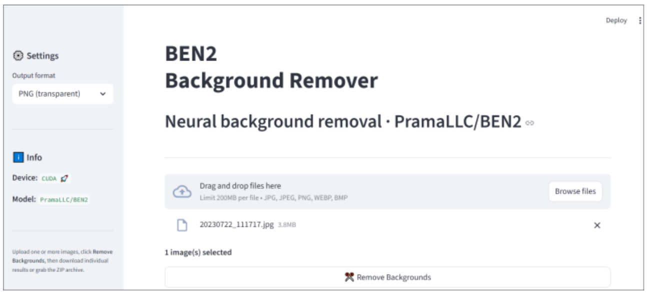

Fig. 4. Graphical user interface (GUI) for single and batch image uploading

The preprocessing module performs image normalization, resolution change, and data preparation for feeding into the neural network. This ensures the stability of the model regardless of the input image parameters.

The segmentation module (the core of the system) implements the inference of the BEN2 model, which includes the initial segmentation of the object; determination of border regions; refinement of boundaries using boundary enhancement mechanisms.

The results visualization module provides the user with the ability to compare the original and processed images (before/after mode), and also displays additional quality metrics, including: object accuracy coefficient; boundary clarity indicator (Fig. 5).

For this example (Fig. 5), the object accuracy coefficient is 77 %, and the edge sharpness index is 59.1, which indicates a fairly good sharpness of the image edges in the image with the background removed.

The results export module allows you to save processed images in PNG format with transparency support or as an archive (ZIP) for batch processing.

A demonstration of batch processing is shown in Fig. 6 and Fig. 7.

Conclusions. Based on the results of the study, the following conclusions can be drawn:

A thorough analysis of existing approaches to automatic background removal has been conducted – from classical threshold segmentation methods to modern deep learning-based architectures such as U-Net and ViT.

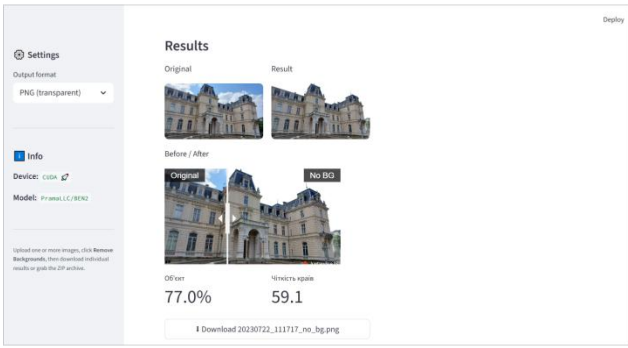

Fig. 5. Segmentation results visualization: “Before/After” comparison and quality metrics

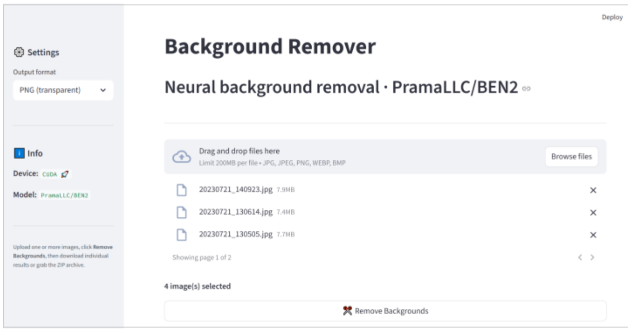

Fig. 6. Batch processing interface for multiple multimedia content files

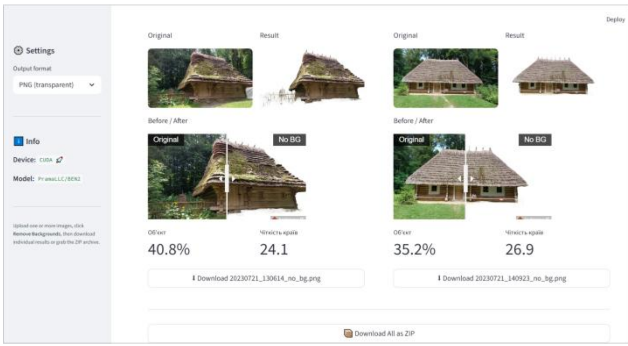

Fig. 7. Examples of automatic background removal for complex multimedia objects

The hybrid neural network architecture BEN2, which implements a two-stage approach: coarse segmentation and further refinement of boundaries using the CGM pipeline, is justified and applied.

The process of separating objects is mathematically modeled through the alpha matting equation, which allows processing translucent transitions and complex contours.

A software application based on the Streamlit framework has been developed, which integrates the BEN2 model and provides a user-friendly interface.

As a result of the research, a system was created that provides highly accurate segmentation of objects even in difficult conditions (complex background texture, uneven lighting); batch image processing functionality was implemented, which increases the efficiency of working with large volumes of multimedia content; high-quality visualization of results was achieved using interactive elements (before/after sliders) and providing metrics of accuracy and clarity of boundaries, and the feasibility of using the proposed approach for applied areas such as e-commerce, graphic design, and media was confirmed.

Directions for further work may include: expanding the post-processing functionality, such as implementing additional feathering and edge filtering methods to completely eliminate background pixel remnants in ultra-complex scenes; optimization for mobile devices; adapting the BEN2 architecture to work on devices with limited computing resources without significant loss of accuracy; working with video content, e.g. developing an algorithm for automatic background removal in a video stream in real time, which is relevant for streaming and video conferencing, and integrating with generative AI, developing modules for automatically replacing the removed background with generative images that match the style of the target object.

Declaration on the use of generative AI. During the preparation of this work, the authors used ChatGPT and Grammarly for grammar and spell checking, as well as for rephrasing and reformulating the text.

References

1. Weingerl P., Hladnik A., Dejana Javoršek D. Development of a machine learning model for extracting image prominent colors. *Color Research & Application*. 2020. No. 45(3), pp. 409–426. DOI: https://doi.org/10.1002/col.22485.
2. Gonzalez R. C., Woods R. E. *Digital Image Processing. 4th ed*. New York: Pearson, 2018. 1019 p.
3. Szeliski R. Computer Vision: Algorithms and Applications (2nd Edition). Springer, 2022. 1232 p.
4. Kovalenko S., Kovalenko S., Mikhnova O., Kovalenko A., Pelikh D., Severin V., "An Approach to Blood Cell Classification Based on Object Segmentation and Machine Learning," *2023 IEEE 4th KhPI Week on Advanced Technology (KhPIWeek)*, Kharkiv, 2023, pp. 1–6. DOI: 10.1109/KhPIWeek61412.2023.10312903.
5. Chen J., Lu Y., Yu Q., Luo X., Adeli E., Wang Y., Lu L., Yuille A. L., Zhou Y. *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. arXiv, preprint arXiv:2102.04306, 2021. 13 p. DOI: 10.48550/arXiv.2102.04306.
6. Kovalenko S. M., Kutsenko O. S., Kovalenko S. V., Kovalenko A. S. Approach to the automatic creation of an annotated dataset for the detection, localization and classification of blood cells in an image. *Radio Electronics, Computer Science, Control*. 2024, no. 1, pp. 128–139. DOI: 10.15588/1607-3274-2024-1-12.
7. Kovalenko A., Severyn V. Optimization of the architecture and hyperparameters of the U-Net model to improve the quality of biological objects segmentation. *Journal of the Ternopil National Technical University*. 2025, vol 120, no. 4, pp. 21–31. DOI: https://doi.org/10.33108/visnyk_tntu2025.04.021.
8. Dosovitskiy, A. An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New York City, 2021, pp. 45–67.
9. Khanorkar, A., Pawar, B., Singh, D., Dhanbhar, K., Mangrulkar, N. Background removal of video in realtime. *3C TIC. Cuadernos de desarrollo aplicados a las TIC*. 2022, vol. 11, no. 2, pp. 195–206. DOI: 10.17993/3ctic.2022.112.195-206.
10. Meyer M., Spruyt J. *Ben: Using confidence-guided matting for dichotomous image segmentation*. arXiv preprint arXiv:2501.06230, 2025. DOI: 10.48550/arXiv.2501.06230.

11. *Meet BEN2, the AI Background Eraser That Actually Works And It Is Open Source.* DigiAlps LTD. Available at: https://digialps.com/meet-ben2-the-ai-background-eraser-that-actually-works-and-it-is-open-source/ (accessed 24.02.2026).
12. Qin X., Dai H., Hu X., Fan D. P., Shao L., Van Gool L. Highly accurate dichotomous image segmentation. *European Conference on Computer Vision.* Cham, Springer Nature Switzerland. 2022, pp. 38–56. DOI: 10.1007/978-3-031-19797-0_3.
13. *BEN2: Background Erase Network*. Available at: https://huggingface.co/PramaLLC/BEN2 (accessed 24.02.2026).
14. Khorasani M., Abdou M., Fernández J. H. *Streamlit for Web Development*: *Build and Scale Secure Python-Powered Apps with Streamlit*. Apress Berkeley, CA, 2025. XXV, 413 p. DOI: https://doi.org/10.1007/979-8-8688-1826-4.

Received 05.03.2026
Accepted 02.04.2026
Published 20.05.2026

УДК 004.932:004.8

С. В. КОВАЛЕНКО, кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», професор кафедри системного аналізу та інформаційно-аналітичних технологій, м. Харків, Україна; e-mail: serhii.kovalenko@khpi.edu.ua; ORCID: https://orcid.org/0000-0001-8763-0862
С. М. КОВАЛЕНКО, кандидатка технічних наук, доцентка, Національний технічний університет «Харківський політехнічний інститут», доцентка кафедри програмної інженерії та інтелектуальних технологій управління, м. Харків, Україна; e-mail: svitlana.kovalenko@khpi.edu.ua; ORCID: https://orcid.org/0000-0001-6770-6778
Т. Є. АЛЕКСАНДРОВА, докторка технічних наук, професорка, Національний технічний університет «Харківський політехнічний інститут», завідувачка кафедри системного аналізу та інформаційно-аналітичних технологій, м. Харків, Україна; e-mail: tetyana.aleksandrova @khpi.edu.ua; ORCID: https://orcid.org/ 0000-0001-9596-0669
М. М. МАЛЬКО, кандидат технічних наук, доцент, Національний технічний університет «Харківський політехнічний інститут», професор кафедри системного аналізу та інформаційно-аналітичних технологій, м. Харків, Україна; e-mail: maxim.malko@khpi.edu.ua; ORCID: https://orcid.org/ 0000-0002-0125-2141

АВТОМАТИЗОВАНА СЕГМЕНТАЦІЯ ОБ’ЄКТІВ ТА ВИДАЛЕННЯ ФОНУ В СИСТЕМАХ ОБРОБКИ МУЛЬТИМЕДІЙНОГО КОНТЕНТУ

В сучасних умовах стрімкого розвитку цифрових технологій обробка зображень посідає ключове місце в галузях комп’ютерного зору та графіки. Однією з найбільш затребуваних задач є автоматичне відокремлення об’єктів переднього плану від фону, що має критичне значення для вебдизайну, електронної комерції, рекламної індустрії та систем доповненої реальності. Традиційні методи ручного редагування є трудомісткими та залежними від досвіду користувача, особливо при роботі зі складними текстурами або нечіткими межами об’єктів. Розвиток методів глибокого навчання дозволяє автоматизувати ці процеси, забезпечуючи високу точність та швидкість обробки в реальному часі. Метою роботи є аналіз існуючих алгоритмів автоматичного видалення фону та розробка ефективної методології сегментації об’єктів із використанням сучасних моделей комп’ютерного зору для підвищення якості мультимедійного контенту. У роботі застосовано комплексний підхід, що базується на використанні гібридної нейромережевої архітектури BEN2. Ключовою особливістю методу є впровадження інноваційного конвеєра Confidence Guided Matting, який реалізує двоетапний байєсівський підхід: спочатку виконується груба сегментація (BEN Base) для створення «чернетки» маски, а потім прицільне уточнення меж у зонах низької впевненості моделі (BEN Refiner). У межах дослідження розроблено програмний застосунок на базі фреймворку Streamlit, який забезпечує автоматизований інференс моделі BEN2. Система підтримує завантаження зображень у популярних графічних форматах, пакетну обробку даних та візуалізацію результатів за допомогою інтерактивного слайдера. Експериментально підтверджено високу ефективність алгоритму: для складних об’єктів архітектура забезпечує точне матування волосся та дрібних деталей, мінімізуючи артефакти типу «рваних країв». Процес обробки оптимізовано шляхом використання локального кешування моделей та підтримки прискорення на GPU/CPU. Запропонований підхід на основі архітектури BEN2 та механізмів уточнення меж продемонстрував вищу точність порівняно з класичними одноетапними методами сегментації. Розроблена система є масштабованою та придатною для інтеграції в реальні інформаційні системи обробки мультимедіа, що дозволяє значно скоротити час підготовки графічного контенту та підвищити його професійний рівень.

Ключові слова: обробка мультимедійного контенту, автоматичне видалення фону, сегментація зображень, глибоке навчання, альфа-матування, архітектура BEN2.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Коваленко Сергій Володимирович / Kovalenko Serhii Volodymyrovych
Автор 2 / Author 2: Коваленко Світлана Миколаївна / Kovalenko Svitlana Mykolayivna
Автор 3 / Author 3: Александрова Тетяна Євгенівна / Aleksandrova Tetyana Evgenivna
Автор 4 / Author 4: Малько Максим Миколайович / Malko Maksym Mykolayovych