EFFECTIVENESS RESEARCH OF USING VIT NEURAL NETWORK ARCHITECTURE FOR CLASSIFYING THE DESTROYED BUILDINGS REMAINS

Hladun Oleksandr
Bachelor student
Molchanova Maryna
Teacher
Zalutska Olha
Teacher
Mazurets Oleksandr
Ph.D in Engineering Science, Associate Professor
Computer Science Department
Khmelnytskyi National University, Ukraine

In the current global landscape, numerous regions are grappling with the aftermath of military conflicts, technological catastrophes, and natural disasters. These events have led to the extensive destruction of both civilian and critical infrastructure, resulting in a pressing need for effective methods to assess and analyze the remnants of damaged buildings. Traditional assessment techniques often fall short in such scenarios due to safety concerns, accessibility issues, and the sheer scale of destruction [1]. To address these challenges, there is an increasing reliance on advanced technologies capable of rapidly collecting, processing, and interpreting large volumes of photographic data. Robotic ground systems have emerged as invaluable tools in this context, offering the ability to navigate hazardous or hard-to-reach areas without endangering human lives. Equipped with high-resolution imaging capabilities, these robots can capture detailed visuals of damaged structures, providing critical data for analysis.

The integration of artificial intelligence, particularly deep learning techniques, has further enhanced the efficiency of damage assessment processes. Convolutional Neural Networks (CNNs), a subset of deep learning models, have demonstrated remarkable proficiency in image recognition tasks [2]. Their ability to learn hierarchical features from visual data makes them particularly suited for identifying and classifying various forms of structural damage. For instance, CNNs can differentiate between types of materials, detect specific damage patterns, and assess the severity of destruction with high accuracy.

Recent studies have highlighted the effectiveness of these models in post-disaster scenarios. For example, research utilizing CNN architectures has achieved significant success in classifying building damage levels from aerial and satellite imagery [3, 4]. These models can process vast datasets to identify patterns indicative of structural compromise, facilitating timely and informed decision-making in disaster response efforts [5]. Despite these advancements, challenges remain in the realm of damage classification [6]. Many existing models are limited by the scope of their training datasets, often encompassing a narrow range of damage types or structural variations. This limitation underscores the necessity for expanding training datasets to include a broader spectrum of building types, materials, and damage scenarios [7]. Enhancing the diversity and volume of training data will improve model generalization, enabling more accurate assessments across varied contexts.

In conclusion, the deployment of robotic ground systems combined with sophisticated deep learning models represents a promising approach to the complex task of assessing damaged infrastructure. By leveraging these technologies, it is possible to conduct rapid, accurate, and safe evaluations of structural damage, thereby supporting effective disaster response and recovery operations. Continued research and development in this field are essential to overcome existing limitations and to refine these tools for broader application in diverse disaster scenarios.

The purpose of the work is effectiveness research of using ViT neural network architecture for classifying the destroyed buildings remains.

To achieve the goal, software was designed and developed that implements neural network analysis of images of the remains of destroyed buildings. Photographic images from cameras installed on robotic equipment that works directly on construction or dismantling sites will be used as input data.

The input data is the ViT neural network for training, a formed dataset with images of elements of the remains of destroyed buildings. The input of the training stage is a set of images of the remains of destroyed buildings, which are used to train the ViT model. At the first stage, input images are prepared in accordance with the requirements of each neural network architecture. For the MobileNetV3 model, this includes resizing images to 224×224 pixels, converting them to tensors, and normalizing values in the range from -1 to 1.

The second stage is to further train the models, evaluate their performance on key accuracy metrics, and save the optimal options for further integration.

As part of the second part of the software development, the third stage involves creating a software application with the integration of trained models. The fourth stage involves testing this application in conditions close to real-world use to verify its stability and accuracy of image analysis.

The output data is a pre-trained ViT model capable of performing automated analysis of photofixation of the remains of destroyed buildings with the results presented in the form of a short analytical report.

Figure 1 shows a diagram of the interaction of the system modules. The software for neural network analysis of photo data of the remains of destroyed buildings consists of four main modules. The input image preparation module is responsible for pre-processing the data before transferring it to the neural network models. In particular, the image is resized to 224×224 pixels, converted to a tensor, and normalized by tensor values. The neural network model training module implements the model training process using the following parameters: number of epochs – 6, batch size – 64. Training is performed on a pre-prepared set of construction debris images. The software application for using neural networks provides interactive user interaction with pre-trained neural networks, in particular, it allows you to view photo images, edit the image database, and perform neural network analysis of photos to classify types of construction debris.

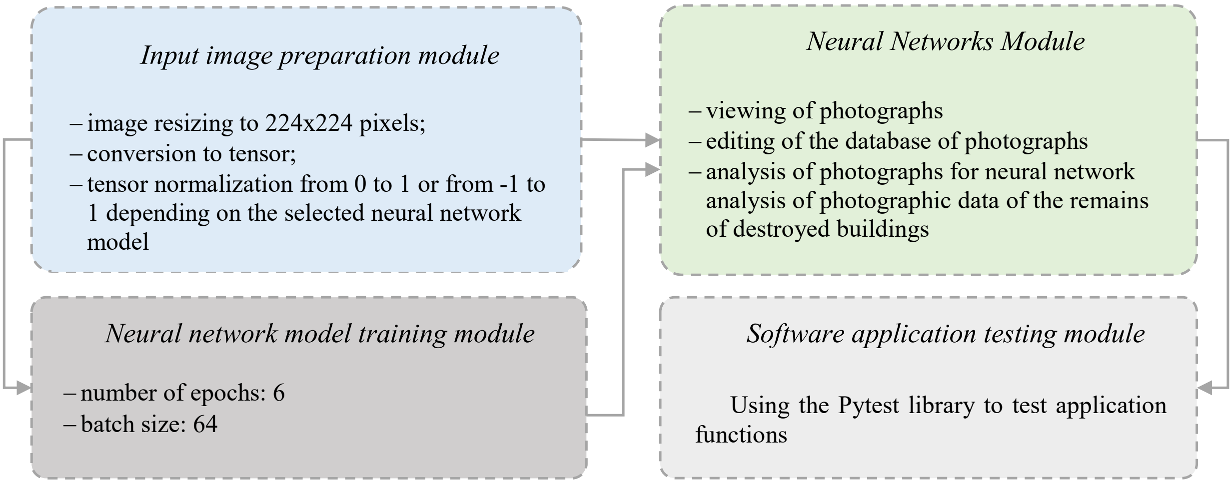

Figure 1. Diagram of interaction of software modules for ViT neural network analysis of photo data of remains of destroyed buildings

The training results (Table 1) of the ViT neural network model demonstrate high efficiency in classifying images of the destroyed buildings remains.

Table 1 – Results of training of MobileNetV3neural network model

Garbage category	Precision	Recall	F1-score
		Micrometrics	
Brick	0.98	0.98	0.98
Concrete	0.99	0.99	0.99
Foam	0.98	0.99	0.98
General_w	0.91	0.92	0.91
Gypsum_board	0.99	0.98	0.99
Pipes	0.88	0.92	0.90
Plastic	0.90	0.83	0.86
Stone	1.00	0.99	0.99
Tile	0.98	0.97	0.97
Wood	0.98	1.00	0.99
		Macrometrics	
Accuracy			0.97
Macro avg	0.96	0.96	0.96
Weighted avg	0.97	0.97	0.97

Training of neural network models implements the process of learning models using the following parameters: number of epochs – 6, batch size – 64. Training is performed on a pre-prepared set of images of construction debris. The software application for using neural networks provides interactive user interaction with pre-trained neural networks, in particular, it allows viewing photo images, editing the image database, and performing neural network analysis of photos to classify types of construction debris. First, the input images are prepared according to the requirements of the neural network architecture. For the ViT model, this includes resizing the images to 224×224 pixels, converting them to tensors, and normalizing the values in the range from -1 to 1.

The graph (Fig. 2) shows the ROC curves for the ViT multi-class construction waste classification model. The ROC curve for each class shows the ratio between the level of false positives and the level of true positives. The area under the curve (AUC) is a metric that characterizes the ability of the model to separate one class from others. In the graph, all curves show AUC values close to 1.00 for each of the presented classes: brick, concrete, foam, general_w, gypsum_board, pipes, plastic, stone, tile, wood. This indicates that the model classifies objects without errors within the test sample. The curves pass through the point (0, 1), which corresponds to the zero level of false positives and the maximum level of true positives. The presence of such results indicates the full discriminatory ability of model in the specified sample.

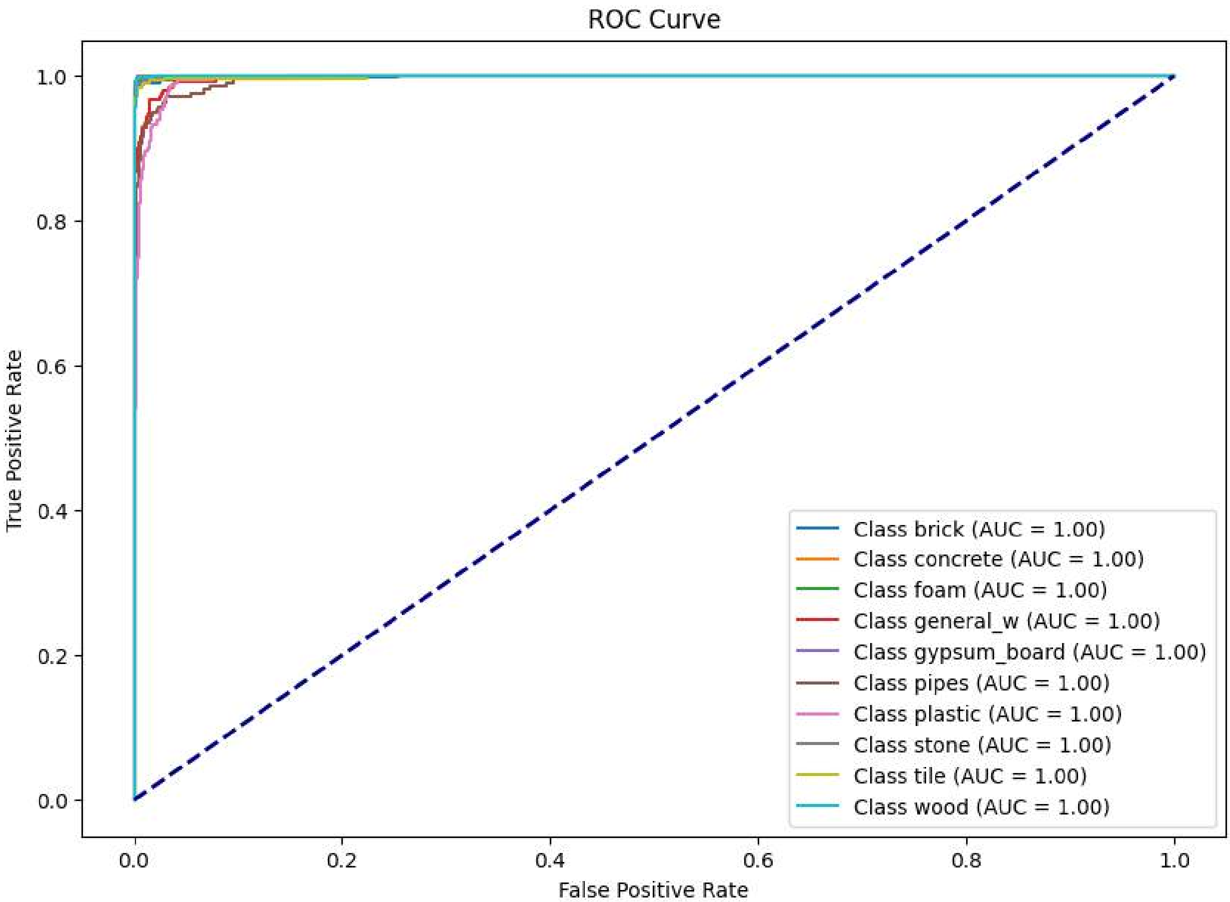

Figure 2. ROC curves for the ViT neural network model

Obtained results demonstrate the effectiveness of ViT as an architecture for multi-class object classification tasks in the field of construction debris image processing. To finally assess the suitability of the model for practical application, it is necessary to additionally consider the Precision-Recall metrics, error analysis, and stability on independent test samples. The results obtained demonstrate the potential of implementing such solutions in practical monitoring and management systems after disasters. Integration of neural network models into robotic systems significantly accelerates the process of identifying and processing remains in the destruction zones, which is relevant in the context of modern challenges. Thus, the developed solution has both scientific and applied significance, opening up prospects for further development in the field of automated analysis of critical situations.

References

1. Liu, C., Sui, H., & Huang, L. (2020). Identification of building damage from UAV-based photogrammetric point clouds using supervoxel segmentation and latent Dirichlet allocation model. Sensors, 20(22), 6499.
2. Kharysh I., Sobko O., & Mazurets O. (2024). Designing CNN Neural Network Model for Detecting Fractures of Lower Extremities by X-ray Images. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. Dubrovnik, Croatia, 91-96.
3. Mazurets O. V., Klimenko V. I., Molchanova M. O., & Sultanov A. V. (2024). Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones. Global Science: Prospects and Innovations. Proceedings of the 10th International scientific and practical conference. Cognum Publishing House. Liverpool, United Kingdom, 198-207.
4. Pokhytun, A., Mazurets, O., Molchanova, M., & Tyschenko, O. (2024). Method for Neural Network Detecting Changed Images of People's Faces Using CNN. New Horizons in Scientific Research: Challenges and Solutions. Proceedings of the 1st International Scientific and Practical Conference, 35-40.
5. Novak, Y., & Mazurets, O. (2023). Practical Application of Method of Automated Personal Identification by Fingerprints Using Convolution Neural Networks. Proceedings of V International Scientific and Practical Conference «Modern strategies of global scientific solutions». 2023. Stockholm, Sweden, International Scientific Unity, 136-140.
6. Mazurets, O., Zalutska, O., Tyschenko, O., & Bohdanova, A. (2024). An Approach to Using MobileNet CNN-model for Gesture Recognition. Proceedings of XXIII International Scientific and Practical Conference «Problems of Science and Technology: the Search for Innovative Solutions», 59-64.
7. Zharnovskyi, O., Mazurets, O., & Sobko, O. (2024). Approach to Identification of Artificial Intelligence-Generated People Images by Means of Machine Learning. Key Aspects of the Development of Scientific Research in Modern Conditions. Proceedings of the XLV International scientific and practical conference, 69-73.