ISSN 1814-4225 (print)
Radioelectronic and Computer Systems, 2025, no. 2(114) ISSN 2663-2012 (online)

UDC 004.932:004.4’2 **doi: 10.32620/reks.2025.2.04**

Vitalii SERDECHNYI, Olesia BARKOVSKA, Andriy KOVALENKO, Anton HAVRASHENKO, Vitalii MARTOVYTSKYI

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

RESEARCH ON MACHINE LEARNING METHODS FOR DETECTING OBJECTS IN DIFFICULT SHOOTING CONDITIONS

*The **subject matter** of the article is research into machine learning methods for object detection in images and videos under complex urban conditions, particularly under poor lighting, the presence of precipitation, high scene complexity, and limited computational resources. The **goal** of this research is to identify the most effective deep learning models based on convolutional neural networks for object detection tasks under challenging imaging conditions, considering the practical requirements for accuracy and processing speed. The **tasks** to be solved are: analysis of object detectors (YOLO v8–11, DETR, SSD, Mask R-CNN, Faster R-CNN, RetinaNet); preparation of a dataset with real weather conditions and pedestrian environments in Ukraine; experimental evaluation of selected detectors using the metrics mAP@0.5, mAP@.5:.95, Recall, Precision, IoU, FPS, and F1-Score; and analysis of the obtained results. The **methods** used are: convolutional neural networks, automated image annotation, comparative analysis of quality metrics (F1-score, mAP@0.5:.95, Precision, Recall, IoU, FPS), and manual correction of annotations. The following **results** were obtained: the YOLOv10-m and YOLOv11-m models demonstrated the best quality indicators under conditions of limited visibility and varying lighting. The YOLOv11-m model was the most balanced in terms of accuracy and speed across all tested conditions - snow, rain, and sunshine. YOLOv11-m is recommended as the baseline model for implementation in real-time systems, particularly in intelligent assistants for people with visual impairments. **Conclusions**: The scientific novelty of the results obtained is as follows: 1) a comprehensive evaluation of modern deep learning architectures for object detection (YOLOv8–v11, Faster R-CNN, SSD, Mask R-CNN, DETR, RetinaNet) was carried out under non-laboratory conditions, including real weather scenarios such as snow, rain, and poor lighting, which are typical for urban environments in Eastern Europe; 2) the software tool for automated model evaluation was developed, allowing simultaneous testing of multiple architectures and visualization of performance metrics (F1-score, mAP@0.5, mAP@.5:.95, IoU, Precision, Recall, FPS) with support for manual annotation correction and comparative model analysis; 3) it was experimentally established that the YOLOv11-m model demonstrates the best balance of accuracy and inference speed across various complex imaging conditions, justifying its recommendation as a baseline model for real-time vision-based assistive systems.*

.

Keywords: *method; detection; image; object; video; YOLO; weather conditions; model.*

1. Introduction

Despite the active development of deep learning approaches for road object detection, most existing studies have been conducted under controlled or laboratory conditions that do not reflect the complexity of real-world environments. In particular, many state-of-the-art solutions have been designed for use in countries with well-structured road infrastructure, consistent markings, predictable traffic behavior, and relatively homogeneous weather conditions. In contrast, Ukrainian urban environments are significantly more variable, featuring non-standard infrastructure, inconsistent or missing road markings, a high density of vulnerable road users, and frequent adverse weather conditions. Given the absence of a universal architecture that would simultaneously ensure high accuracy, robustness against environmental factors, and real-time performance, this work focuses on experimentally evaluating and comparing modern object detection models using a custom dataset. The dataset includes real-world video sequences captured under various weather and times of day in Ukrainian cities. This allows for the identification of architectures that provide the best trade-off between detection accuracy, processing speed, and adaptability to practical road monitoring tasks.

1.1 Motivation

The development of intelligent assistance systems for people with visual impairments involves the implementation of some tasks. These include the design, research, and improvement of methods that ensure the operation of the system’s modules. The main goal of such a system is to analyze information about the surrounding environment and to provide support to people with visual

impairments during movement. The implementation of these tasks is possible by using computer vision methods and devices that accompany the user in the following sequence:

- development of a method to determine environmental conditions (lighting, weather conditions, etc.) using highly heterogeneous data (images, data from LiDAR sensors, audio data from microphones);
- improvement of an adaptive method (considering environmental conditions) for preprocessing data from cameras and LiDAR sensors;
- improvement of a method for object detection and tracking in images or videos based on data from cameras and LiDAR sensors;
- improvement of a method for predicting the trajectories of dynamic objects with consideration of environmental conditions;
- study and application of audio analysis methods aimed at improving the detection of environmental conditions. This includes the development of a high-precision voice interface for data input, clarification of obstacles, route correction, and conducting dialogue to obtain necessary information;
- development of a prototype device to assist people with visual impairments in everyday life.

After studying the factors affecting environmental conditions and the safety, comfort, mobility, and independence of users, the following groups were identified:

- obstacles and hazards in the environment;
- difficulties in mobility;
- lack of access to information;
- social interaction and risk of criminal assaults;
- transportation-related dangers;
- dependence on external assistance.

Particular attention in the research is given to groups 1–3: obstacles and hazards in the environment (obstructions on the path such as uneven surfaces, stairs, uncovered manholes, curbs, and objects on the road can cause injuries; lack of accessible navigation tools; encounters with stray or wild animals); mobility difficulties (people with visual impairments often face challenges when crossing roads, using public transportation, or even navigating familiar places due to the lack of clear understanding of their surroundings; absence of tactile markers or auditory signals); lack of access to information (absence of visual information or failure to consider the needs of people with visual impairments can lead to unawareness of potential threats or dangers).

1.2 State of the art

Since the accuracy of identifying objects in the surrounding environment is significantly affected by weather conditions, which can complicate object detection, this study includes a review of research that specifically focused on weather condition recognition as one of the preparatory steps for detecting moving objects.

The review [1] considered almost all common types of weather phenomena that negatively affect the ability of sensors to perceive and measure data, including rain, snow, fog, haze, strong light, and contamination. In addition, it presented datasets, simulators, and experimental tools with weather-condition support. This review was conducted in the context of applying weather recognition methods to automated driving systems (ADS). The authors of [1] assessed the impact of each type of adverse weather condition (light rain <4 mm/h, heavy rain >25 mm/h, dense smoke/fog vis<0.1 km, fog vis<0.5 km, haze/smoke height>2 km, heavy snow, temperature) on the main types of sensors (LiDARs, radars, ground-penetrating radar, cameras, stereo cameras) (Table 1). The intensity of the impact is evaluated on a scale ranging from 0 to 5, where:

0 – negligible: effects that can be almost ignored;

1 – minor: effects that rarely cause detection errors;

2 – slight: effects that cause minor errors in special cases;

3 – moderate: effects that cause perception errors up to 30% of the time;

4 – severe: effects that cause perception errors in more than 30% but less than 50% of cases;

5 – critical: noise or obstruction leading to false detection or detection failure.

Thus, LiDAR is affected by heavy snow and fog, which significantly limits its effectiveness. Radar is the most resistant to weather conditions and is almost unaffected by rain and fog. Cameras are vulnerable to bright light and contamination, which can completely block their perception.

Table 1

Impact of weather conditions on sensors in automated driving systems [1]

Weather condition	LiDAR	Radar	Camera	Thermal imager
Light rain (< 4 mm/h)	2	0	3	2
Heavy rain (> 25 mm/h)	3	1	4	3
Fog (visibility < 0.5 km)	4	0	4	1
Snow	5	2	2	2
Bright light	2	2	5	2
Sensor contamination	3	2	5	4

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Thermal imagers operate reliably at night but have limitations when contaminated or at low temperatures.

To compensate for the weaknesses of individual sensors, various sensor combinations were used, resulting in improved object detection accuracy under challenging weather conditions. The "Camera + LiDAR + Radar" configuration provides the highest reliability in snow and fog. Some of the main sensor fusion configurations in a single system are summarized in Table 2.

This study also demonstrates the use of machine learning algorithms and image processing methods - De-raining, HDDM+ (fog removal), De-noising (WeatherNet), and point cloud segmentation.

Table 2

Sensor fusion configuration [1]

Sensor configuration	Target weather condition	Comment
Radar + LiDAR	Rain	Combines the advantages of radar accuracy and LiDAR range
Camera + LiDAR (SLS-Fusion)	Fog	Improves object detection in fog
RGB Camera + Thermal imager	Night, Snow	Thermal imagers are effective at night and in snow
Camera + LiDAR + Radar	Snow, Rain, Fog	The most reliable configuration

The study [2] proposed a weather phenomenon classification model based on a deep convolutional neural network called MeteCNN. The classification accuracy of the MeteCNN model reached 92.68%, which demonstrates competitive classification performance among some of the main models (such as VGG16, ResNet34, EfficientNet-B7) on the dataset created by the authors - WEAPD (Weather Phenomenon Database).

The weather phenomenon image database, constructed according to meteorological criteria, contains 6,877 images for 11 types of weather phenomena, featuring complex and noisy backgrounds, with each image including additional objects that create interference. The best recognized categories are: rainbow, lightning, and frost with accuracy of 100%, 99%, and 98%, respectively. The lowest performing categories are "snow" and "glaze," each with an accuracy of 85% (Table 3).

The results presented in Table 3 and Figure 1 were obtained using the following hyperparameters and settings: 13 convolutional layers (conv layers) and 6 pooling layers, batch size - 16, initial learning rate - 0.001, learning rate decay applied, optimizer - stochastic gradient descent (SGD) with a momentum of 0.9.

Table 3

Quality evaluation metrics of the optimized architecture of the MeteCNN neural network model [2]

Category	Precision	Recall	F1-Measure
Hail	0.98	0.98	0.98
Rainbow	1.00	1.00	1.00
Snow	0.85	0.85	0.85
Rain	0.91	0.96	0.93
Lightning	1.00	0.97	0.99
Dew	0.97	0.99	0.98
Sandstorm	0.93	0.99	0.96
Frost	0.93	0.85	0.89
Smog/Fog	0.98	0.94	0.96
Rime	0.88	0.89	0.88
Glaze	0.85	0.84	0.85
Average value	0.93	0.93	0.93

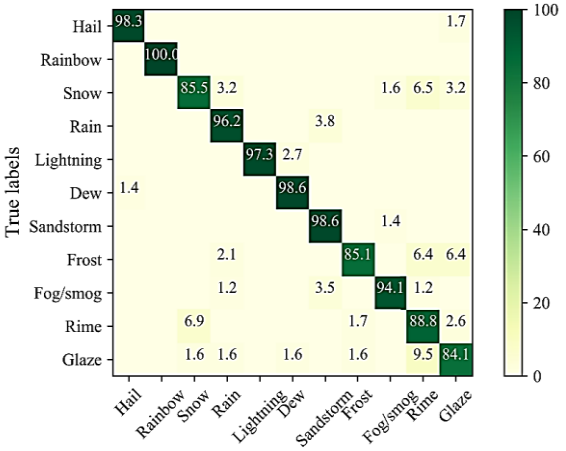

Fig. 1. Predicted labels [2]

A drawback of the proposed model is that it confuses certain categories of weather phenomena, which may be attributed to the similarity and complexity of the images.

Among the other studies considered, one that stands out because of its significant achievements and relevance to the related problem area is the study [3]. The researchers proposed a new approach, IA-YOLO (Image-Adaptive YOLO), to improve object detection under adverse weather conditions – in fog and low-light environments.

The main distinction of the proposed approach from existing ones is the introduction of an additional image processing module, which enables the extraction of hidden useful information by removing weather-specific content. The proposed DIP (digital image processing) module consists of six differentiated filters, namely: Defog, White Balance (WB), Gamma correction, Contrast, Tone adjustment, and Sharpen.

The qualitative indicators from the previous study [3] demonstrate the object detection results using

YOLOv3 II (columns 1, 3, and 5) and the proposed IA-YOLO (columns 2, 4, and 6) on synthetic images from VOC_Foggy_test (top row) and real fog images from the RTTS foggy dataset (bottom row). The proposed method learns to reduce fogginess and sharpen image edges, which ensures better detection performance with fewer missed and false detections.

The results shown in Figure 2 confirm that the proposed image enhancement steps for foggy or low-light conditions provide an advantage for IA-YOLO over various YOLOv3 detector modifications across different datasets.

Study [4] is dedicated to the creation of the multimodal OLIMP dataset, which was designed for training artificial intelligence models to perceive the surrounding environment. It includes four types of data (modalities):

- images;
- ultra-wideband radar signatures;
- narrow-band data streams;
- acoustic data.

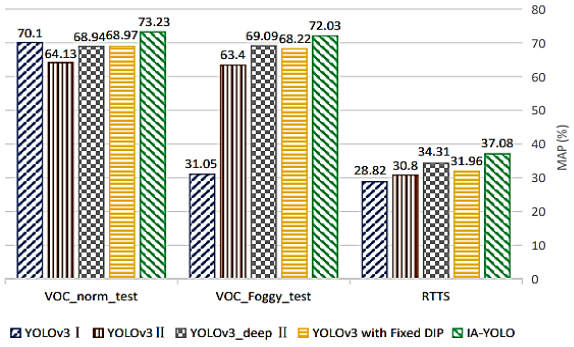

Fig. 2. Quantitative performance indicators of the study [4]

A distinctive feature of the dataset is that it includes signals from ultra-wideband radar and acoustic sensors, and it is primarily focused on dense urban traffic situations.

To demonstrate the effectiveness of the prepared dataset, experiments were conducted, which demonstrated that the use of individual modalities yields significantly lower accuracy in determining environmental conditions, indicating the necessity of expanding input data modalities by utilizing various sensors for data acquisition (Table 4). The obtained results confirmed that combining multiple data sources (datasets of RGB images, UWB Radar data, and Acoustic data into a dataset named Fusion) helped reduce errors and improve detection quality.

The article [5], we present a detailed comparative analysis of various early fusion techniques for combining visible and thermal images to enhance object detection using convolutional neural networks. The authors demonstrate that static fusion methods, particularly channel summation and concatenation, significantly improve CNN performance under low-light and thermally complex scenarios. This research is relevant for computer vision tasks in nighttime surveillance and can be applied to autonomous monitoring systems.

The review article [6] focuses on methods for fusing LiDAR data with other sources (e.g., optical, radar) to improve the accuracy of forest attribute estimation. While the primary application is environmental monitoring, this paper systematizes multimodal data fusion approaches that are applicable to object detection tasks in complex scene geometry or poor visibility situations. The article is especially relevant for those exploring LiDAR integration into object detection systems in urban or road environments.

Several recent studies have focused on object detection under complex environmental conditions, emphasizing the growing need for robust computer vision systems for real-world scenarios. In the work by Chan et al. [7], the authors proposed a non-machine-learning-based system for detecting preceding vehicles under various lighting and weather conditions. The proposed method combined four vehicular structure-related visual cues with a particle filter to improve detection stability. This study demonstrated that when carefully designed, traditional model-based approaches can still be competitive under non-ideal visual conditions.

Table 4

Evaluation results of object detection accuracy (pedestrian, cyclist, vehicle, tram) using different training data modalities: ImageNet (for MobileNet-v2), UWB Radar, Acoustic, and Fusion. The table presents key metrics: Precision (P), Recall (R) [4]

Object	MobileNet (P), %	MobileNet (R), %	MobileNet (AP), %	UWB Radar (P), %	UWB Radar (R), %	Acoustic (P), %	Acoustic (R), %	Fusion (P), %	Fusion (R), %
Pedestrian	84	54	53	46	36	20	17	86	54
Cyclist	77	70	67	45	52	44	15	81	69
Vehicle	81	48	47	8	0	40	38	82	48
Tram	86	76	75	0	0	61	64	90	76

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Table 5

Comparative table of video file analyzer performances

YOLOvX [11]	
Features	One-stage approach to object detection, fast processing due to simple architecture. The model is implemented based on a convolutional neural network with single-shot prediction, optimized for real-time operation.
Weaknesses	Low accuracy for small objects, sensitivity to lighting.
Application	Real-time detection, simultaneous identification of multiple objects.
Faster R-CNN [12]	
Features	Two-stage detection, high accuracy for complex scenes. The model is implemented based on a Region Proposal Network (RPN) for generating regions of interest, with VGG or ResNet as the backbone architecture.
Weaknesses	Slow processing, high computational complexity.
Application	High-precision applications: medical diagnostics, security systems.
SSD [11, 13]	
Features	One-stage approach, a compromise between speed and accuracy. Utilizes multiple scales for predicting objects of different sizes.
Weaknesses	Low accuracy for small objects, limited adaptability.
Application	Pedestrian detection, vehicle monitoring in surveillance systems.
Mask R-CNN [14]	
Features	Two-stage approach: first, region detection, then classification and segmentation using detailed object masks. The model is implemented based on a Region Proposal Network (RPN) for object localization.
Weaknesses	Slow processing, high computational requirements.
Application	Segmentation: medical diagnostics, object labeling.
DETR [15]	
Features	Based on transformer architecture for object detection in a scene.
Weaknesses	Slow processing, high resource requirements.
Application	Research, general-purpose object detection tasks.
RetinaNet [11]	
Features	Based on transformer architecture for object detection in a scene.
Weaknesses	Slow processing, high resource requirements.
Application	Research, general-purpose object detection tasks.

However, the absence of learning-based adaptability limits the scalability of the model to highly dynamic environments.

Chellappa et al. [8] explored the fusion of acoustic and visual sensors for vehicle tracking to address visibility issues caused by poor weather or occlusions. Their system integrates data from multiple modalities to improve detection accuracy, especially in cases where visual input alone may be unreliable. The research highlights the value of multi-sensor fusion; however, it also notes challenges in synchronizing and calibrating heterogeneous sensor data streams for real-time applications.

In a more recent deep-learning-based approach, Ghosh [9] proposed an enhancement to Faster R-CNN by incorporating several region proposal networks (RPNs) of varying sizes. This modification allows the detector to capture objects of different scales more effectively, particularly in adverse weather conditions such as blizzards, snowfalls, and wet road conditions. The system was evaluated on three public datasets (DAWN, CDNet 2014, and LISA) and achieved notable average precision improvements (up to 95.16%), outperforming conventional single-RPN architectures. The results of this study underscore the benefit of architectural modification for improving robustness in vehicle detection tasks.

Finally, the work by Tumas et al. [10] introduced the ZUT dataset, which includes thermal imaging and weather annotations for pedestrian detection in low-visibility scenarios. Their experiments revealed that the existing datasets lack adequate environmental variability and often suffer from insufficient thermal resolution. By using a modified YOLOv3 on 16-bit thermal data, their system achieved up to 89.1% mAP, confirming the potential of sensor-specific datasets to improve detection under fog, snow, or rain. This study contributes not only a valuable dataset and a benchmark for testing detection systems under real-world environmental constraints.

It is worth noting that modern computer vision systems focused on object detection in complex environments use various types of annotations, which significantly affect the accuracy of model training and the effectiveness of detection under real-world conditions:

- ROI (Region of Interest);
- AOI (Area of Interest);
- POI (Point of Interest);
- Bounding Box;
- Polygonal Segmentation;
- Keypoints Annotation;
- Line Annotations;
- Semantic Segmentation;
- Object Tracking;

3D Bounding Boxes.The correct choice of data annotation format determines not only the efficiency of the training process but also the flexibility of further model adaptation to new tasks.

Based on the analysis of existing solutions, the most common types of annotations used in object recognition tasks can be identified: ROI (Region of Interest), AOI (Area of Interest), POI (Point of Interest), classical rectangular bounding boxes, polygonal segmentation, keypoints annotation, line annotations, semantic segmentation, object tracking in video streams, and 3D bounding boxes.

In most cases, for the task of object detection on roads - both in static images and in video streams—annotations of the Bounding Box, Semantic Segmentation, or Object Tracking types are used. This is explained by the balance between object positioning accuracy and the relative ease of generating such annotations by the chosen model.

Before selecting a model, we reviewed the existing neural network analyzers designed for detecting moving objects in video sequences: YOLO (You Only Look Once), Faster R-CNN (Region-based Convolutional Neural Network), SSD (Single Shot Multibox Detector), Mask R-CNN, DETR (DEtection TRansformer), RetinaNet (Table 5). Based on the analysis of modern object detection models, including YOLOv11, Faster R-CNN, SSD, Mask R-CNN, DETR, and RetinaNet, it can be concluded that one-stage models, such as YOLOv11 and SSD, provide high processing speed, which is critically important for real-time tasks; however, they are inferior to two-stage approaches in terms of accuracy when recognizing small or partially occluded objects. In contrast, models based on the Region Proposal Network (such as Faster R-CNN, Mask R-CNN) demonstrate higher accuracy but require significantly more computational resources and exhibit lower frame rates (FPS). In addition, next-generation models built on transformers (DETR, RetinaNet) demonstrate potential for task universalization in detection; however, they are characterized by the highest resource requirements and slow processing speeds.

Therefore, the problem arises of insufficient accuracy of existing obstacle recognition systems under real operating conditions in Ukraine. This makes the development of an adaptive road obstacle recognition system highly relevant, one capable of operating on images obtained in complex and heterogeneous urban conditions of Ukrainian cities and demonstrating improved accuracy through targeted fine-tuning of the model on localized datasets.

1.3 Aims and tasks of the work

The aim of this work was to investigate machine learning methods based on convolutional neural networks for object detection under challenging imaging conditions, such as poor lighting, precipitation, and a large number of scene objects, considering the limited resources of the video recorder.

To achieve the stated goal, the following tasks must be accomplished:

- to analyze object detectors (YOLO v8–11, DETR, SSD, Mask R-CNN, Faster R-CNN, RetinaNet);
- to prepare a dataset with real weather conditions and pedestrian environments in Ukraine;
- to conduct an experimental study of the selected detectors using the metrics mAP@0.5, mAP@.5:.95, Recall, Precision, IoU, FPS, F1-Score;
- to analyze the obtained results.

Further research will involve the implementation of the proposed method and testing on real data using fusion-sensor inputs.

The proposed approach and experimental results can be used to create intelligent assistance systems for people with visual impairments, autonomous driving systems, and urban navigation systems [16-17].

The structure of the paper is designed to systematically investigate and evaluate the effectiveness of modern deep learning methods for object detection under complex visual conditions. In Section 1. Introduction, the paper outlines the relevance of the problem in the context of urban infrastructure in Ukraine, highlighting the problems of object detection under poor lighting, precipitation, and non-standard conditions. This is further confirmed in Section 1.1 Motivation, which emphasizes the need for assistive systems for people with visual impairments and discusses in detail practical issues related to navigation and obstacle avoidance in real-world conditions. Section 1.2 State of the art provides a detailed review of the existing literature on environmental recognition and object detection in adverse scenarios, analyzing approaches such as IA-YOLO and MeteCNN, and presenting the importance of multisensor data fusion and contextual adaptation. Section 1.3 Aims and tasks of the work defines the main objectives of the research, namely

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

the experimental comparison of object detectors such as YOLO v8–11, DETR, SSD, Mask R-CNN, Faster R-CNN, and RetinaNet on real data obtained in Ukrainian cities, using metrics such as mAP@0.5, mAP@.5:.95, IoU, Precision, Recall, FPS, and F1-Score. The experimental process is given in Section 2.2 Materials and methods of research, which describe the collection of the dataset, annotation procedures, and development of a special software tool for testing models and visualization of metrics. This section also describes in detail the conditions under which the testing was performed, including snow, rain, sunny weather, as well as different levels of illumination and scene complexity. The evaluation results are discussed in Section 3. Results discussion, which presents a comparative analysis of the performance of all tested models, identifying YOLOv10-m and YOLOv11-m as the most efficient architectures under different conditions. Finally, Section 4 Conclusions summarizes the key findings, highlights the suitability of YOLOv11-m as a baseline model for real-time vision-based systems, and outlines future directions, including the integration of LiDAR and audio inputs, to improve detection reliability in complex environments. The paper concludes with declarations regarding conflicts of interest, funding sources, data availability, and a comprehensive literature section that confirms the scientific validity of the study.

2. Materials and methods of research

Based on the identified environmental conditions affecting the safety of visually impaired pedestrians, the first task was to develop a method for determining environmental conditions - lighting, weather conditions, etc. - using highly heterogeneous data such as images, LiDAR sensor data, and audio data from microphones) will be addressed by evaluating the quality of obstacle detection using existing artificial intelligence-based models (Figure 3).

This will make it possible to identify the weaknesses of models in detecting vehicles, people, animals, stairs, open manholes, and general obstacles for further improvement and optimization of methods under specific imaging conditions (precipitation, fog, complex scenes) and for certain object categories.

The diagram illustrates the workflow of developing and improving a neural network model for analyzing objects in real-world environments. The five stages of this process are described below.

Dataset collection under real environmental conditions. In this stage, video files of various real-world objects and environmental conditions (vehicles, people, traffic lights, etc.). The task of this stage is to provide the model with real test data corresponding to the actual conditions in which the developed intelligent assistant will record video with the following parameters:

- constant camera movement during walking;
- footage captured by medium-accuracy cameras;
- various weather conditions;
- different lighting conditions;
- varying scene complexity;
- a variable number of objects in the frame.

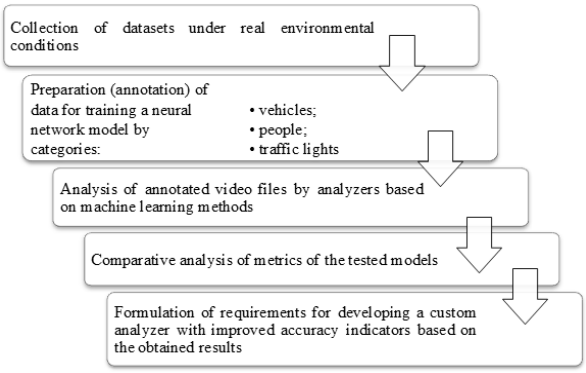

Fig. 3. Research workflow diagram to justify the need to improve the accuracy and speed of object detection under various imaging conditions

Data collection was conducted in different environments (urban, natural), considering changes in lighting, weather conditions, and shooting perspectives, to create a representative dataset that ensured effective model training.

The statistical indicators of the data prepared for training the model under real-world conditions are as follows (Fig. 4):

- the duration of video files that took part in testing the neural network models varied from 40 to 480 seconds;
- dataset characteristics: number of files - 5 files, including 3 in cloudy weather, 2 in snowy weather, 1 in rainy weather, 2 in sunny weather, and 1 in twilight weather.

Ім'я	Дата	Тип	Розмір	Довжина
Video #1 (City, Day, Cloudy, Rain)	16.12.2024 13:14	Файл MP4	1 210 638 КБ	00:08:08
Video #2 (City, Early Morning, Snow)	21.11.2024 7:11	Файл MP4	374 822 КБ	00:02:31
Video #3 (City Day, Cloudy, Snow)	21.11.2024 15:50	Файл MP4	827 889 КБ	00:05:34
Video #4 (City, Day, Sunny)	01.05.2025 16:08	Файл MP4	100 493 КБ	00:00:52
Video #5 (City, Day, Sunny)	23.04.2025 16:08	Файл MP4	76 693 КБ	00:00:39

Fig. 4. Video files for test dataset formation

The preparation (annotation) of data for testing the neural network model is necessary to define the categories, sizes, and locations of objects in the frame that the neural network model must recognize.

The identified categories are vehicles, people, and traffic lights (Figure 5).

To obtain a dataset to test machine learning methods under real-world conditions, video files were defragmented into frames at a rate of 30 frames per second. As a result, we obtained a test dataset containing 31,920 frames.

Fig. 5. Sample frames used for testing

Since the test videos in this experiment were collected using only video cameras, without considering distance estimation to objects, the chosen annotation type was 2D bounding boxes, which conveniently represent the coordinates of an object within the frame and provide effective neural network training. 2D bounding boxes do not contain depth information, which is important for determining distance. This form of annotation is among the most commonly used in computer vision tasks for training neural networks, particularly for models such as YOLO or SSD, where each object is defined by its bounding box coordinates and class.

To automate the research process, a software tool was developed that enables rapid video annotation, automatically initiated by the YOLOv11 model, followed by manual correction and the ability to evaluate the performance of a wide range of models (YOLOv8–11, Faster R-CNN, Mask R-CNN, SSD, DETR, RetinaNet).

The research tasks include the creation of a user-friendly interface for viewing and editing bounding boxes automatically annotated by the YOLO model, integration of recognition models, performance analysis using metrics such as F1-score, mAP, IoU, Precision, Recall, and FPS, as well as result visualization. Figure 6 shows that each detected object (e.g., vehicles and pedestrians) is marked with a bounding box and a corresponding class label.

Figure 7 contains a fragment of a text file containing the detection results generated by the model during inference (prediction). Each line corresponds to one of the objects found in the image and contains the following data: object class (e.g., car or person), the model's confidence score, and bounding box coordinates in a normalized format – values [x_center, y_center, width, height], usually relative to the image size. These coordinates are used for result visualization, but specifically during model performance evaluation, such as when calculating mAP or other quality metrics.

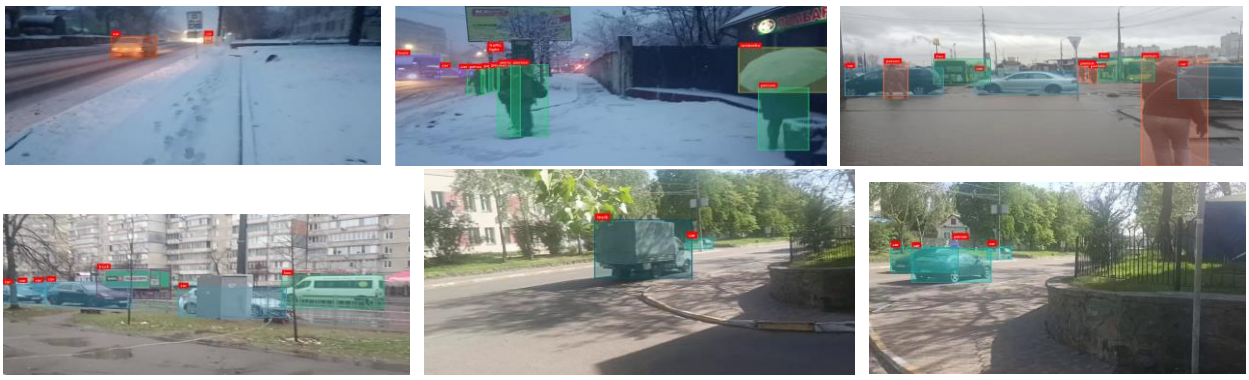

Fig. 6. Examples of corrected video frame annotations
used to test the aforementioned neural network models (ground truth annotation)

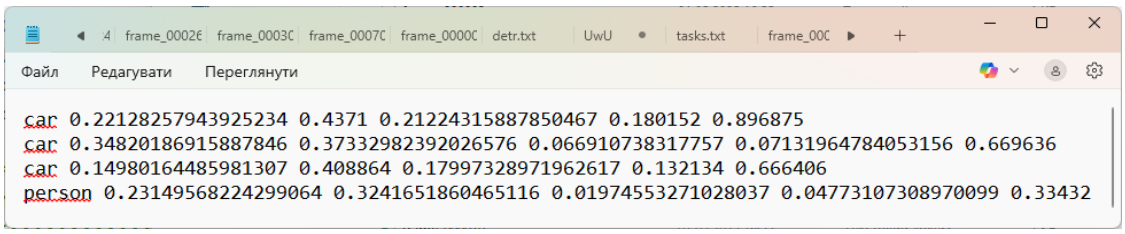

Fig. 7. The image annotation results are suitable for further analysis and comparison with ground truth data

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

The next stage involves the analysis of the video files annotated by the detector based on AI methods. The task of this stage is to ensure the verification of annotation quality and preliminary testing of the speed and efficiency of existing models, as well as to identify the strengths and weaknesses of the selected model in the specific task for their further improvement prior to implementation in a hardware-software solution for assisting visually impaired individuals on city streets. The comparative analysis demonstrates the model's effectiveness under different recording conditions for detecting various objects, such as snowy, rainy, overcast weather in the morning or during evening twilight (Tables 5–6).

All studied models are shown in Figure 8 – YOLO, SSD, Mask R-CNN, Faster R-CNN, DETR, RetinaNet.

The developed software tool input frames to each model. Each model performs annotation, after which the annotation is compared with the ground truth, and the key metrics are calculated. Using the example of the YOLOv11 model, the annotation results for all frames shown in Figures 8 and 9 are presented below.

3. Results discussion

The test results of the models under snowy, overcast weather early in the morning are shown in Table 6.

Based on the results presented in Table 6, the best models in terms of F1-score were YOLOv10-b (0.55), YOLOv10-m (0.54), and YOLOv11-x (0.54). They demonstrate a balance between precision and recall. The fastest models according to the research results are YOLOv8-n, YOLOv10-n, and YOLOv10-s; however, their F1-score is noticeably lower (up to 0.48), making them suitable for tasks where speed is critical rather than maximum accuracy.

The YOLOv11-m model demonstrated the most accurate object positioning (according to the IoU metric), but its F1-score was average (0.51), making it best suited for tasks in which precise object localization is crucial. The highest precision according to the precision metric was obtained by the SSD model (0.75), but it had a low Recall (0.14), indicating many missed objects. According to the mAP@.5:.95 metric, all models showed extremely low results (up to 0.06), which can be explained by the challenging conditions (overcast, snowy weather in the morning). The YOLOv10-m model demonstrated the best overall performance according to this metric.

Thus, in snowy, overcast weather early in the morning, the best-performing model across all metrics was YOLOv10-m, which achieved a high F1-score (0.54), the highest mAP@.5:.95 (0.06), high FPS (34), sufficiently high IoU (0.92), and balanced Precision (0.66) and Recall (0.45). An alternative with a higher F1-score is YOLOv10-b (0.55); however, this model has slightly lower metrics in other areas. If speed is more critical, at the cost of reduced accuracy, good results were obtained by YOLOv10-n (FPS = 49, F1 = 0.35) or YOLOv8-n (FPS = 56, F1 = 0.41).

The test results of the models under overcast and rainy weather conditions during the day are presented in Table 7.

Based on the results presented in Table 7, the leaders in object detection quality in this study were YOLOv10-l and YOLOv11-m. These models demonstrated the highest F1-score values (0.77–0.78) and mAP@.5:.95 (0.53–0.59) with acceptable FPS (11–15), making them optimal for tasks where recognition quality in challenging weather conditions is critical. A compromise between speed and quality is offered by the YOLOv8-s, YOLOv11-s, and YOLOv8-m models. They provide higher performance (FPS) with moderate quality (0.36–0.38 mAP@.5:.95). The YOLOv8-n and YOLOv11-n models exhibit the highest processing speed (26–32 FPS) but at the expense of recognition quality. These models are suitable for preliminary selection or tracking. The Non-YOLO models (RetinaNet, Faster R-CNN, Mask R-CNN, DETR) demonstrated poorer performance in terms of both speed and quality.

Thus, in overcast, rainy weather during the day, the best model overall was YOLOv11-m, which showed a high F1-score (0.78), mAP@.5:.95 (0.47), sufficiently high FPS (15), high IoU (0.91), and balanced Precision (0.79) and Recall (0.76).

The test results of the models under sunny weather during the day are presented in Table 8.

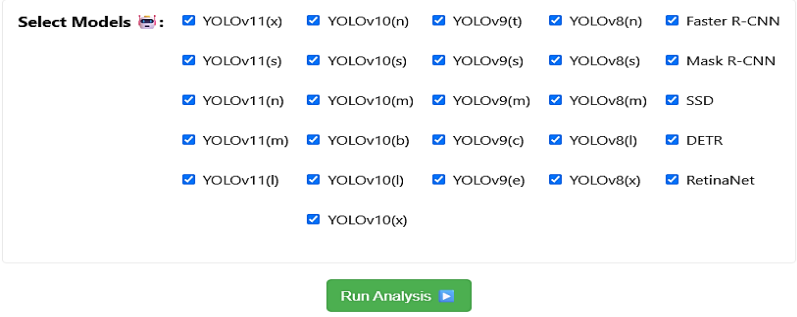

Fig. 8. Studied neural network models

Table 6

Research results of analyzers in snowy, overcast weather early in the morning

Model	F1-score	FPS	IoU	Precision	Recall	mAP@.5:.95	mAP@0.5
YOLO_10_m	0.54	34.00	0.92	0.66	0.45	0.06	0.07
YOLO_10_l	0.53	23.00	0.91	0.61	0.47	0.05	0.06
YOLO_9_e	0.52	12.00	0.89	0.50	0.54	0.05	0.06
YOLO_10_b	0.55	27.00	0.91	0.64	0.48	0.04	0.05
YOLO_11_l	0.53	27.00	0.90	0.62	0.45	0.04	0.05
YOLO_8_m	0.47	30.00	0.90	0.47	0.47	0.04	0.05
YOLO_9_c	0.50	25.00	0.90	0.53	0.48	0.04	0.05
RETINA_net	0.42	4.00	0.85	0.57	0.34	0.02	0.04
YOLO_10_n	0.35	49.00	0.91	0.50	0.27	0.03	0.04
YOLO_10_s	0.48	48.00	0.91	0.62	0.39	0.04	0.04
YOLO_10_x	0.49	16.00	0.91	0.49	0.50	0.04	0.04
YOLO_11_m	0.51	30.00	0.96	0.62	0.43	0.03	0.04
YOLO_11_s	0.45	45.00	0.89	0.47	0.44	0.03	0.04
YOLO_11_x	0.54	13.00	0.91	0.61	0.48	0.04	0.04
YOLO_8_l	0.52	18.00	0.89	0.53	0.51	0.03	0.04
YOLO_8_s	0.44	50.00	0.90	0.44	0.44	0.04	0.04
YOLO_8_x	0.50	12.00	0.89	0.48	0.52	0.03	0.04
YOLO_9_m	0.51	31.00	0.89	0.52	0.51	0.04	0.04
YOLO_9_s	0.44	28.00	0.90	0.48	0.41	0.03	0.04
SSD	0.23	14.00	0.87	0.75	0.14	0.02	0.03
YOLO_11_n	0.39	43.00	0.88	0.43	0.36	0.02	0.03
YOLO_8_n	0.41	56.00	0.89	0.54	0.33	0.02	0.03
YOLO_9_t	0.33	28.00	0.89	0.34	0.33	0.02	0.03
FASTER_RCNN	0.31	4.00	0.81	0.25	0.43	0.01	0.02
MASK_RCNN	0.38	4.00	0.81	0.31	0.48	0.01	0.02

Table 7

Research results of analyzers under overcast, rainy weather during the day

Model	F1-score	FPS	IoU	Precision	Recall	mAP@.5:.95	mAP@0.5
YOLO_10_l	0.77	11.00	0.86	0.81	0.74	0.47	0.59
YOLO_11_m	0.78	15.00	0.91	0.79	0.76	0.47	0.53
YOLO_10_b	0.72	13.00	0.88	0.76	0.69	0.40	0.48
YOLO_10_x	0.78	8.00	0.85	0.80	0.77	0.31	0.42
YOLO_8_l	0.77	10.00	0.86	0.76	0.77	0.33	0.42
YOLO_11_n	0.55	26.00	0.84	0.61	0.51	0.31	0.41
YOLO_8_x	0.76	7.00	0.85	0.73	0.79	0.32	0.40
YOLO_9_m	0.71	13.00	0.86	0.71	0.71	0.32	0.40
YOLO_11_s	0.68	23.00	0.85	0.66	0.69	0.33	0.39
YOLO_10_n	0.55	23.00	0.86	0.67	0.47	0.30	0.38
YOLO_8_s	0.68	28.00	0.82	0.67	0.70	0.27	0.38
RETINA_net	0.62	4.00	0.82	0.74	0.54	0.27	0.37
YOLO_9_c	0.72	12.00	0.85	0.70	0.75	0.29	0.37
YOLO_8_m	0.73	15.00	0.86	0.70	0.76	0.29	0.36
YOLO_11_l	0.74	12.00	0.87	0.72	0.75	0.27	0.34
YOLO_11_x	0.74	7.00	0.85	0.72	0.77	0.26	0.34
YOLO_10_m	0.70	15.00	0.87	0.72	0.68	0.26	0.33
YOLO_9_s	0.65	12.00	0.87	0.69	0.62	0.26	0.33
YOLO_9_e	0.74	6.00	0.86	0.70	0.79	0.25	0.32

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Continuation of the Table 7

Model	F1-score	FPS	IoU	Precision	Recall	mAP@.5:.95	mAP@0.5
YOLO_10_s	0.64	15.00	0.86	0.68	0.59	0.26	0.31
YOLO_8_n	0.58	32.00	0.83	0.62	0.55	0.23	0.30
FASTER_RCNN	0.66	4.00	0.80	0.57	0.77	0.20	0.28
MASK_RCNN	0.62	4.00	0.81	0.54	0.75	0.19	0.26
YOLO_9_t	0.59	13.00	0.84	0.68	0.53	0.20	0.26
DETR	0.51	22.00	0.79	0.41	0.66	0.15	0.25

Table 8

Research results of analyzers in sunny weather

Model	F1-score	FPS	IoU	Precision	Recall	mAP@.5:.95	mAP@0.5
YOLO_11_m	0.87	16.00	0.93	0.87	0.87	0.76	0.83
YOLO_8_s	0.83	27.00	0.88	0.86	0.80	0.59	0.76
YOLO_9_c	0.86	13.00	0.90	0.88	0.83	0.53	0.59
YOLO_10_s	0.84	23.00	0.89	0.91	0.78	0.48	0.58
YOLO_9_m	0.86	14.00	0.89	0.85	0.87	0.46	0.56
YOLO_8_l	0.86	11.00	0.90	0.87	0.85	0.48	0.55
YOLO_11_x	0.86	5.00	0.90	0.85	0.87	0.43	0.52
RETINA_net	0.84	4.00	0.89	0.93	0.76	0.41	0.51
YOLO_10_l	0.82	11.00	0.90	0.84	0.80	0.40	0.49
FASTER_RCNN	0.70	4.00	0.85	0.58	0.89	0.34	0.47
MASK_rcnn	0.74	4.00	0.86	0.64	0.87	0.37	0.47
YOLO_8_x	0.82	8.00	0.90	0.78	0.87	0.38	0.47
YOLO_9_e	0.90	7.00	0.90	0.92	0.87	0.39	0.47
YOLO_9_t	0.79	15.00	0.89	0.90	0.70	0.38	0.45
YOLO_10_m	0.80	16.00	0.91	0.85	0.76	0.41	0.45
YOLO_8_n	0.77	37.00	0.89	0.84	0.70	0.38	0.44
YOLO_9_s	0.84	13.00	0.89	0.86	0.81	0.37	0.42
YOLO_11_s	0.77	23.00	0.88	0.75	0.80	0.33	0.40
YOLO_10_x	0.83	9.00	0.90	0.83	0.83	0.36	0.39
YOLO_8_m	0.81	16.00	0.89	0.80	0.81	0.28	0.35
YOLO_10_b	0.83	15.00	0.91	0.86	0.80	0.32	0.35
YOLO_10_n	0.71	22.00	0.92	0.85	0.61	0.32	0.35
SSD	0.62	12.00	0.88	1.00	0.44	0.27	0.34
YOLO_11_l	0.79	13.00	0.91	0.77	0.81	0.28	0.32
YOLO_11_n	0.73	27.00	0.90	0.80	0.67	0.27	0.31

Based on the results presented in Table 8, the model with the best overall performance in this study is YOLOv11-m. It demonstrated the highest F1-score (0.87), the highest mAP@.5:.95 (0.76), a suitable real-time processing speed (16 FPS), a high localization level with an IoU of 0.93, and the best balance between Precision and Recall (0.87 for both metrics). A good trade-off between speed and quality was also observed for the YOLOv8-s (F1 = 0.83, mAP@.5:.95 = 0.59, FPS = 27), YOLOv10-s (F1 = 0.84, mAP@.5:.95 = 0.48, FPS = 23), and YOLOv9-c (F1 = 0.86, mAP@.5:.95 = 0.53, FPS = 13) models. These models demonstrated a good level of accuracy at higher speeds. The YOLOv8-n (FPS 37) and YOLOv11-n (FPS 27) models were the fastest among all models but exhibited lower accuracy (mAP@.5:.95 ≈ 0.27–0.38). The models built on RetinaNet, Faster R-CNN, Mask R-CNN, and SSD showed weaker results in most metrics: low FPS (around 4), moderate F1-score (0.62–0.74), and low mAP@.5:.95 (0.27–0.41).

Thus, under sunny weather conditions, the best model based on all the indicators, as in the previous study, was YOLOv11-m.

4. Conclusions

This study focuses on the evaluation and analysis of the performance of a wide range of neural network models (YOLOv8-11, Faster R-CNN, Mask R-CNN, SSD,

DETR, RetinaNet) in the task of detecting moving objects in video sequences under challenging recording conditions, such as poor lighting, precipitation, and a large number of scene objects. To automate this process and enable detailed analysis, a specialized software tool was developed that allows rapid video annotation, annotation correction, and evaluation of model performance based on key metrics: F1-score, mAP@0.5, mAP@.5:.95, IoU, Precision, Recall, and FPS.

Within the framework of the assigned tasks:

- an overview of modern object detectors (YOLOv8–11, DETR, SSD, Mask R-CNN, Faster R-CNN, RetinaNet) was conducted with an analysis of their architectural features, advantages, and disadvantages in the context of road obstacle recognition;
- a dataset of 31,920 files was prepared, reflecting real weather conditions of Ukrainian cities, in particular - the city of Obukhov and the city of Kremenchuk (sunny, rainy, snowy weather, various lighting conditions), as well as typical pedestrian conditions. The dataset was annotated using Bounding Box annotations;
- a software tool was developed to enable the automated testing and evaluation of models using key metrics: F1-score, mAP@0.5, mAP@0.5:.95, IoU, Precision, Recall, FPS;
- experimental testing of YOLOv8–11, DETR, SSD, Mask R-CNN, Faster R-CNN, and RetinaNet models was conducted on real video fragments collected in urban environments of Ukraine under various weather conditions, which allowed us to identify the dependence of model accuracy and performance on weather conditions and time of day.

The analysis results showed that under challenging conditions (snow, dusk, rain), the YOLOv10-m and YOLOv11-m models demonstrated the best balance between accuracy and speed. In particular, YOLOv10-m achieved the highest F1-score in snowy weather, while YOLOv11-m achieved the highest F1-score under rainy and sunny conditions. High-FPS models, such as YOLOv10-n and YOLOv8-n, are suitable for scenarios in which high speed is critical but maximum accuracy is not required. The two-stage models (Faster R-CNN, Mask R-CNN) and transformer-based models (DETR) are inferior to YOLO in terms of speed and flexibility, making the latter more suitable for mobile solutions.

Thus, the YOLOv11-m model demonstrated the highest stability across all recording conditions and can be recommended as a baseline model for further development of real-time object recognition systems, particularly intelligent assistance systems for visually impaired individuals.

The scientific significance of this study lies in the comprehensive evaluation of state-of-the-art convolutional neural network architectures (YOLOv8–11, Faster R-CNN, SSD, Mask R-CNN, DETR, RetinaNet) under non-ideal environmental conditions, including low light, precipitation, and complex urban scenes. This study contributes to the development of deep learning-based object detection for solving real-world problems specific to the infrastructure of Eastern European cities. The comparative analysis of models using several performance metrics (F1-score, mAP@0.5, mAP@0.5:.95, IoU, Precision, Recall, FPS) under different weather scenarios improves the understanding of the robustness and adaptability of detection algorithms and paves the way for the development of context-aware perception systems.

The practical significance of this research lies in the further implementation of the results in real-time assistive technologies for people with visual impairments, autonomous driving systems, and intelligent solutions for urban monitoring. The proposed methodology allows for the adaptation and optimization of detection systems for deployment in heterogeneous, dynamic, and visually complex environments typical of Ukrainian infrastructure.

Future research involves the integration of fusion data (LiDAR, audio, RGB) to improve the reliability of the system under limited visibility conditions.

Contribution of authors: Formulation of the problem **Olesia Barkovska. Andriy Kovalenko**; hardware platform investigation and development tools selection **Vitalii Serdechnyi, Anton Havrashenko**. Methodology of the experiments **Olesia Barkovska**; ANN models realization **Vitalii Serdechnyi, Anton Havrashenko**; analysis and processing of the obtained results **Olesia Barkovska, Vitalii Martovytskyi**; finalization of the draft article version **Olesia Barkovsk**a; corrections and postediting **Vitalii Martovytskyi, Andriy Kovalenko**.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship, or otherwise, that could affect the research and its results presented in this paper.

Financing

This work was partially supported under the grant ‘‘An assistant system that accompanies dynamic objects’’ (state registration number of the scientific and technical work: 0124U003837).

Data availability

The manuscript contains no associated data.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence methods while creating the presented work.

All the authors have read and agreed to the published version of this manuscript.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

References

1. Zhang, Y., Carballo, A., Yang, H., & Takeda, K. Perception and sensing for autonomous vehicles under adverse weather conditions: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023, pp. 146-177. DOI: 10.1016/j.isprsjprs.2022.12.021.

2. Xiao, H., Zhang, F., Shen, Z., Wu, K., & Zhang, J. Classification of weather phenomenon from images by using deep convolutional neural network. *Earth and Space Science*, 2021, vol. 8, iss. 5, article no. e2020EA001604. DOI: 10.1029/2020EA001604.

3. Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., & Zhang, L. Image-adaptive YOLO for object detection in adverse weather conditions. *In Proceedings of the AAAI conference on artificial intelligence*, 2022, vol. 36, no. 2, pp. 1792-1800. DOI: 10.48550/arXiv.2112.08088.

4. Mimouna, A., Alouani, I., Ben Khalifa, A., El Hillali, Y., Taleb-Ahmed, A., Menhaj, A., Ouahabi, A., & Ben Amara, N. E. OLIMP: A Heterogeneous Multimodal Dataset for Advanced Environment Perception. *Electronics* 2020, vol. 9, article no. 560. DOI: 10.3390/electronics9040560.

5. Heredia-Aguado, E., Cabrera, J. J., Jiménez, L. M., Valiente, D., & Gil, A. Static Early Fusion Techniques for Visible and Thermal Images to Enhance Convolutional Neural Network Detection: A Performance Analysis. *Remote Sens*. 2025, vol. 17, article no. 1060. DOI: 10.3390/rs17061060.

6. Balestra, M., Marselis, S., Sankey, T. T., Cabo, C., Liang, X., Mokroš, M., & et al. LiDAR data fusion to improve forest attribute estimates: A review. *Current Forestry Reports*, 2024, vol. 10, pp. 281-297. DOI: 10.1007/s40725-024-00223-7.

7. Al-Haija, Q. A., Smadi, M. A., & Zein-Sabatto, S. Multi-Class Weather Classification Using ResNet-18 CNN for Autonomous IoT and CPS Applications. *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2020, pp. 1586-1591. DOI: 10.1109/CSCI51800.2020.00293.

8. Rajib, G. On-road vehicle detection in varying weather conditions using faster R-CNN with several region proposal networks. *Multimedia Tools Appl*, 2021, vol. 80, pp. 25985–25999. DOI: 10.1007/s11042-021-10954-5.

9. Osipov, A., Pleshakova, E., Gataullin, S., Korchagin, S., Ivanov, M., Finogeev, A., & Yadav, V. Deep Learning Method for Recognition and Classification of Images from Video Recorders in Difficult Weather Conditions. *Sustainability*, 2022, vol. 14, article no. 2420. DOI: 10.3390/su14042420.

10. Tumas, P., Nowosielski A., & Serackis, A. Pedestrian Detection in Severe Weather Conditions, *in IEEE Access*, 2020, vol. 8, pp. 62775-62784, DOI: 10.1109/ACCESS.2020.2982539.

11. Tan, L., Huangfu, T., Wu, L., & Chen, W. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC Med Inform Decis Mak*, 2021, vol. 21, article no. 324, DOI: 10.1186/s12911-021-01691-8.

12. Ren, S., He, K., Girshick, R., & Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, vol. 39, no. 6, pp. 1137-1149, DOI: 10.1109/TPAMI.2016.2577031.

13. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. SSD: Single Shot MultiBox Detector. In *Computer Vision–ECCV 2016: 14th European Conference*, 2016, vol. 9905, pp. 21-37, DOI: 10.1007/978-3-319-46448-0_2.

14. Bharati, P., & Pramanik, A. Deep Learning Techniques—R-CNN to Mask R-CNN: A Survey. *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, pp. 657-668. DOI: 10.1007/978-981-13-9042-5_56.

15. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko. S. End-to-End Object Detection with Transformers. In *European conference on computer vision ECCV 2020*, 2020, vol. 12346, pp. 213-229, DOI: 10.1007/978-3-030-58452-8_13.

16. Barkovska, O., & Serdechnyi, V. Intelligent assistance system for people with visual impairments. *Innovative technologies and scientific solutions for industries*, 2024, no. 2(28), pp. 6–16, DOI: 10.30837/2522-9818.2024.28.006.

17. Wu, J., Ye, Y., & Du, J. Multi-objective reinforcement learning for autonomous drone navigation in urban areas with wind zones. *Automation in construction*, 2024, vol. 158, article no. 105253. DOI: 10.1016/j.autcon.2023.105253.

Received 01.03.2025. Accepted 20.05.2025

ДОСЛІДЖЕННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ДЕТЕКТУВАННЯ ОБ'ЄКТІВ У СКЛАДНИХ УМОВАХ ЗЙОМКИ

В. С. Сердечний, О. Ю. Барковська, А. Ю. Коваленко, А. О. Гаврашенко, В. О. Мартовицький

Предметом вивчення в статті є дослідження методів машинного навчання для виявлення об'єктів на зображеннях і відео в складних міських умовах, зокрема за поганого освітлення, наявності опадів, високої

складності сцени та обмежених обчислювальних ресурсів. **Метою** є визначення найбільш ефективних моделей глибокого навчання на основі згорткових нейронних мереж для завдань виявлення об'єктів у складних умовах зйомки з урахуванням практичних вимог до точності та швидкості обробки. **Завдання**: аналіз детекторів об'єктів (YOLO v8–11, DETR, SSD, Mask R-CNN, Faster R-CNN, RetinaNet); підготовку набору даних з реальними погодними умовами та пішохідним середовищем в Україні; експериментальну оцінку обраних детекторів із використанням метрик mAP@0.5, mAP@.5:.95, Recall, Precision, IoU, FPS та F1-Score; аналіз отриманих результатів. Використані **методи**: згорткові нейронні мережі, автоматизоване анотування зображень, порівняльний аналіз метрик якості (F1-score, mAP@0.5:.95, Precision, Recall, IoU, FPS), ручна корекція анотацій. Отримані **результати**: моделі YOLOv10-m і YOLOv11-m показали найкращі показники якості в умовах обмеженої видимості та змінного освітлення. YOLOv11-m виявилась найбільш збалансованою з точки зору точності та швидкості за всіх протестованих умов - сніг, дощ, сонячна погода. Модель YOLOv11-m рекомендована як базова для впровадження в системах реального часу, зокрема в інтелектуальних асистентах для людей з порушенням зору. **Висновки**. Наукова новизна отриманих результатів полягає в наступному: вперше проведено комплексну оцінку сучасних архітектур глибокого навчання для виявлення об'єктів (YOLOv8–v11, Faster R-CNN, SSD, Mask R-CNN, DETR, RetinaNet) в умовах, що не є лабораторними, зокрема за реальних погодних сценаріїв (сніг, дощ, погане освітлення), характерних для міських середовищ Східної Європи; розроблено програмний інструмент для автоматизованої оцінки моделей, що дозволяє одночасно тестувати кілька архітектур і візуалізувати метрики продуктивності (F1-міра, mAP@0.5, mAP@.5:.95, IoU, Precision, Recall, FPS) з підтримкою ручного коригування анотацій і порівняльного аналізу моделей; експериментально встановлено, що модель YOLOv11-m демонструє найкращий баланс між точністю та швидкістю обробки в різних складних умовах зйомки, що обгрунтовує її рекомендацію як базової моделі для систем допомоги в реальному часі на основі комп'ютерного зору.

Ключові слова: метод; виявлення; зображення; об'єкт; відео; YOLO; погодні умови; модель.

Сердечний Віталій Сергійович – асп. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна

Барковська Олеся Юріївна – канд. техн. наук, доц., доц. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна

Коваленко Андрій Анатолійович – д-р техн. наук, проф., проф. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна

Гаврашенко Антон Олегович – асп. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна

Мартовицький Віталій Олександрович – канд. техн. наук, доц., доц. каф. електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна

Vitalii Serdechnyi – PhD Student of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: vitalii.serdechnyi@nure.ua, ORCID: 0009-0007-8828-5803.

Olesia Barkovska – PhD, Associate Professor, Associate Professor at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: olesia.barkovska@nure.ua, ORCID: 0000-0001-7496-4353, Scopus Author ID: 24482907700.

Andriy Kovalenko – Doctor of Technical Sciences, Professor, Professor at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: andriy.kovalenko@nure.ua, ORCID: 0000-0002-2817-9036, Scopus Author ID: 56423229200.

Anton Havrashenko – PhD Student of the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: anton.havrashenko@nure.ua, ORCID: 0000-0002-8802-0529.

Vitalii Martovytskyi – PhD, Associate Professor, Associate Professor at the Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine,
e-mail: vitalii.martovytskyi@nure.ua, ORCID: 0000-0003-2349-0578, Scopus Author ID: 57196940070.