ҚАЗАҚСТАН РЕСПУБЛИКАСЫНЫҢ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РЕСПУБЛИКИ КАЗАХСТАН
MINISTRY OF SCIENCE AND HIGHER EDUCATION OF THE REPUBLIC OF KAZAKHSTAN

ХАЛЫҚАРАЛЫҚ УНИВЕРСИТЕТІ
INTERNATIONAL UNIVERSITY
SINCE 2009
МЕЖДУНАРОДНЫЙ УНИВЕРСИТЕТ

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

2025 (22) 2

сәуір - маусым

ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Халықаралық ақпараттық және коммуникациялық технологиялар журналы
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Меншік иесі: АҚ «Халықаралық ақпараттық технологиялар университеті» (Алматы қ.).
Қазақстан Республикасы Ақпарат және қоғамдық даму министрлігіне мерзімді баспасөз басылымын есепке қою туралы куәлік № KZ82VPY00020475, 20.02.2020 ж. берілген
Тақырып бағыты: ақпараттық технологиялар, ақпараттық қауіпсіздік және коммуникациялық технологиялар, әлеуметтік-экономикалық жүйелерді дамытудағы цифрлық технология.
Мерзімділігі: жылына 4 рет.
Тираж: 100 дана.
Редакция мекенжайы: 050040 Алматы қ., Манас к., 34/1, каб. 709, тел: +7 (727) 244-51-09.
E-mail: ijict@iitu.edu.kz
Журнал сайты: https://journal.iitu.edu.kz

Журнал сайты: https://journal.iitu.edu.kz

Международный журнал информационных и коммуникационных технологий
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)

Собственник: АО «Международный университет информационных технологий» (г. Алматы).

Свидетельство о постановке на учет периодического печатного издания в Министерство информации и общественного развития Республики Казахстан **№ KZ82VPY00020475,** выданное от **20.02.2020 г.**

Тематическая направленность: информационные технологии, информационная безопасность и коммуникационные технологии, цифровые технологии в развитии социо-экономических систем**.**

Периодичность: 4 раза в год.

Тираж: 100 экземпляров.

Адрес редакции: 050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09.

E-mail: ijict@iitu.edu.kz Сайт журнала: https://journal.iitu.edu.kz

«International Journal of Information and Communication Technologies»
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Owner: International Information Technology University JSC (Almaty).
The certificate of registration of a periodical printed publication in the Ministry of Information and Social Development of the Republic of Kazakhstan, Information Committee No. KZ82VPY00020475, issued on 20.02.2020.
Thematic focus: information technology, digital technologies in the development of socio-economic systems, information security and communication technologies
Periodicity: 4 times a year.
Circulation: 100 copies.
Editorial address: 050040. Manas st. 34/1, Almaty. +7 (727) 244-51-09. E-mail: ijict@iitu.edu.kz
Journal website: https://journal.iitu.edu.kz

МАЗМҰНЫ

СОДЕРЖАНИЕ

CONTENTS

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Vol. 6. Is. 2. Number 22 (2025). Pp. 37–56
Journal homepage: https://journal.iitu.edu.kz
https://doi.org/10.54309/IJICT.2025.22.2.003

УДК 004.932:004.93’2:004.89:681.518

DEVELOPMENT AND TESTING OF AN INTELLIGENT MONITORING SYSTEM FOR CONSTRUCTION PROCESSES BASED ON COMPUTER VISION AND MACHINE LEARNING

G.U. Bektemysova[1], A.A. Bykov[1]*, B.A. Nor’ashikin[2], A.A. Keresh[1], M.O. Dauletbek[1]

[1]International Information Technology University, Almaty, Kazakhstan;
[2]Universiti Tenaga Nasional, Kajang, Selangor, Malaysia.
E-mail: a.bykov@edu.iitu.kz

Bektemysova G.U. — Candidate of Technical Sciences, professor, «the Department of Computer Engineering», International Information Technology University, Almaty, Kazakhstan
E-mail: g.bektemisova@iitu.edu.kz, https://orcid.org/0000-0002-0850-0558;
Bykov A.A. — Candidate Technical Sciences, associate professor, «the Department of Computer Engineering», International Information Technology University, Almaty, Kazakhstan
E-mail: a.bykov@edu.iitu.kz, https://orcid.org/0000-0002-9563-5185;
Nor’ashikin B.A. — PhD in Information Systems, associate professor, Deputy Dean (Academic, Quality & Student Affairs), Universiti Tenaga Nasional (UNITEN), Putrajaya, Selangor, Malaysia
E-mail: shikin@uniten.edu.my, https://orcid.org/0000-0002-5653-2482;
Keresh A.A. — Master of Sc., International Information Technology University, Almaty, Kazakhstan
E-mail: arman@keresh.ru, https://orcid.org/0009-0008-9695-0241;
Dauletbek M.O. — Master of Technical Sciences, «the Department of Cybersecurity», International Information Technology University, Almaty, Kazakhstan
E-mail: m.dauletbek@iitu.edu.kz, https://orcid.org/0009-0005-5569-4980.

Abstract. This paper presents the design and testing of an intelligent construction monitoring system using computer vision and machine learning to improve safety and efficiency. The system focuses on automated detection of personal protective equipment (PPE) and safety violations at construction sites. It employs YOLOv11 and ResNet-50 neural networks for real-time object detection and classification. Lightning AI was used for cloud-based data processing. Testing on real-world data showed 70–75 % detection accuracy (mAP@0.5) for workers and PPE, and 92 % accuracy for classifying safety violations. However, under stricter criteria (mAP@0.5:0.95), accuracy dropped to 35 %, highlighting challenges in complex or occluded scenes. The most frequent errors involved helmets and safety glasses misrecognition. The system reduces human error and supports proactive safety management, contributing to lower accident rates and better oversight on construction sites.

Keywords: computer vision, construction monitoring, occupational safety, object detection, deep learning, YOLOv11, ResNet-50, Lightning AI

For citation: G.U. Bektemysova, A.A. Bykov, A.A. Nor'ashikin, A.A. Keresh, M.O. Dauletbek. Development and testing of an intelligent monitoring system for construction processes based on computer vision and machine learning// International journal of information and communication technologies. 2025. Vol. 6. No. 22. Pp. 37–56 (In Eng.). https://doi.org/10.54309/IJICT.2025.22.2.003.

Conflict of interest: The authors declare that there is no conflict of interest.

Funding information. This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant №AP23488873 - The development of an intelligent monitoring system for construction processes based on image recognition models).

КОМПЬЮТЕРЛІК КӨРУ ЖӘНЕ МАШИНАЛЫҚ ОҚЫТУ НЕГІЗІНДЕ ҚҰРЫЛЫС ПРОЦЕСТЕРІН БАҚЫЛАУДЫҢ ИНТЕЛЛЕКТУАЛДЫ ЖҮЙЕСІН ӘЗІРЛЕУ ЖӘНЕ СЫНАУ

Г.У. Бектемісова[1], А.А. Быков[1]*, Б.А. Нурашихин[2], А.А. Кереш[1], М.О. Даулетбек[1]

[1]Халықаралық Ақпараттық Технологиялар Университеті, Алматы, Қазақстан;
[2]Тенага Ұлттық университеті, Каджанг, Селангор, Малайзия.
E-mail: a.bykov@edu.iitu.kz

Бектемісова Г.У. — «Компьютерлік Инженерия» кафедрасының, техника ғылымдарының кандидаты, профессор, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: g.bektemisova@iitu.edu.kz, https://orcid.org/0000-0002-0850-0558;
Быков А.А. — «Компьютерлік Инженерия» кафедрасының, техника ғылымдарының кандидаты, доцент, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: a.bykov@edu.iitu.kz, https://orcid.org/0000-0002-9563-5185;
Нурашихин Б.А. — ақпараттық жүйелер саласындағы PhD докторы, деканның орынбасары, Профессор, Тенага Ұлттық университеті (UNITEN), Каджанг, Селангор, Малайзия
E-mail: shikin@uniten.edu.my, https://orcid.org/0000-0002-5653-2482;
Кереш А.А. — «Компьютерлік Инженерия» кафедрасының, техникалық ғылымдарының магистрі, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: arman@keresh.ru, https://orcid.org/0009-0008-9695-0241;
Даулетбек М.О. — «Киберқауіпсіздік» кафедрасының, техникалық ғылымдарының магистрі, Халықаралық ақпараттық технологиялар университеті, Алматы, Қазақстан
E-mail: m.dauletbek@iitu.edu.kz, https://orcid.org/0009-0005-5569-4980.

Аннотация. Бұл жұмыста құрылыс алаңдарындағы қауіпсіздік пен тиімділікті арттыруға бағытталған интеллектуалды бақылау жүйесі ұсынылады. Жүйе жеке қорғаныс құралдарын (ЖҚҚ) автоматты түрде анықтап, қауіпсіздік ережелерінің бұзылуын тіркейді. YOLOv11 және ResNet-50 нейрондық желілері нысандарды нақты уақытта анықтап, жіктеу үшін қолданылды. Деректерді өңдеу Lightning AI бұлтты платформасы арқылы жүзеге асты. Нақты деректерде сынақ жүргізу нәтижесінде жұмысшылар мен ЖҚҚ анықтау дәлдігі 70–75 % (mAP@0.5) болса, қауіпсіздік бұзушылықтарын жіктеу дәлдігі 92 %-ға жетті. Қатал бағалау кезінде (mAP@0.5:0.95) дәлдік 35 %-ға дейін төмендеді, бұл күрделі көріністер мен жабық нысандарда қиындықтар барын көрсетті. Ең жиі кездесетін қателер – каска мен көзілдірікті дұрыс танымау. Жүйе адам қателігін азайтып, қауіпсіздік бақылауын жақсартады.

Түйін сөздер: computer vision, construction monitoring, occupational safety, object detection, deep learning, YOLOv11, ResNet-50, Lightning AI

Дәйексөздер үшін: Г. У. Бектемісова, А. А. Быков, Б.А. Нурашихин, А. А. Кереш, М. О. Даулетбек. Компьютерлік көру және машиналық оқыту негізінде құрылыс процестерін бақылаудың интеллектуалды жүйесін әзірлеу және сынау//Халықаралық ақпараттық және коммуникалық технологиялар журналы. 2025. Т. 6. No. 22. 37–56бет. (ағылшын тілінде). https://doi.org/10.54309/IJICT.2025.22.2.003.

Мүдделер қақтығысы: Авторлар осы мақалада мүдделер қақтығысы жоқ деп мәлімдейді.

РАЗРАБОТКА И ТЕСТИРОВАНИЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ МОНИТОРИНГА СТРОИТЕЛЬНЫХ ПРОЦЕССОВ НА ОСНОВЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ И МАШИННОГО ОБУЧЕНИЯ

Г.У. Бектемысова[1], А.А. Быков[1]*, Б.А. Нурашихин[2], А.А. Кереш[1], М.О. Даулетбек[1]

[1]Международный университет информационных технологий, Алматы, Казахстан;

[2]Университет Тенага Насиональ (UNITEN), Каджанг, Селангор, Малайзия.

E-mail: a.bykov@edu.iitu.kz

Бектемысова Г.У. — кандидат технических наук, профессор, «Кафедра компьютерной инженерии», Международный университет информационных технологий, Алматы, Казахстан
E-mail: g.bektemisova@iitu.edu.kz, https://orcid.org/0000-0002-0850-0558;

Быков А.А. — кандидат технических наук, ассоциированный профессор, «Кафедра компьютерной инженерии», Международный университет информационных технологий, Алматы, Казахстан
E-mail: a.bykov@edu.iitu.kz, https://orcid.org/0000-0002-9563-5185;

Нурашихин Б.А. — профессор, PhD в области информационных систем, заместитель декана (учебная работа, качество и работа со студентами), Университет Тенага Насиональ (UNITEN), Каджанг, Селангор, Малайзия
E-mail: shikin@uniten.edu.my, https://orcid.org/0000-0002-5653-2482;

Кереш А.А. — магистр технических наук, «Кафедра компьютерной инженерии», Международный университет информационных технологий, Алматы, Казахстан
E-mail: arman@keresh.ru, https://orcid.org/0009-0008-9695-0241;

Даулетбек М.О. — магистр технических наук, «Кафедра кибербезопасности», Международный университет информационных технологий, Алматы, Казахстан
E-mail: m.dauletbek@iitu.edu.kz, https://orcid.org/0009-0005-5569-4980.

Аннотация. В работе представлена интеллектуальная система мониторинга строительных площадок, разработанная с применением компьютерного зрения и машинного обучения для повышения безопасности и эффективности. Основное внимание уделено автоматическому распознаванию средств индивидуальной защиты (СИЗ) и выявлению нарушений техники безопасности. Использовались нейросетевые модели YOLOv11 и ResNet-50 для детекции и классификации объектов. Облачная обработка данных выполнялась с помощью платформы Lightning AI. При испытаниях на реальных данных система показала точность обнаружения работников и СИЗ на уровне 70–75 % (mAP@0.5), а классификация нарушений достигла 92 %. Однако при более строгой оценке (mAP@0.5:0.95) точность снизилась до 35 %, что указывает на сложности в сложных сценах и при частичном перекрытии объектов. Наиболее частыми ошибками стали неверное распознавание касок и защитных очков. Система снижает влияние человеческого фактора и способствует улучшению контроля за соблюдением техники безопасности.

Ключевые слова: компьютерное зрение, мониторинг строительства, охрана труда, обнаружение объектов, глубокое обучение, YOLOv11, ResNet-50, Lightning AI

Для цитирования: Г.У. Бектемысова, А.А. Быков, Б.А. Нурашихин, А.А. Кереш, М.О. Даулетбек. Разработка и тестирование интеллектуальной системы мониторинга строительных процессов на основе компьютерного зрения и машинного обучения // Международный журнал информационных и коммуникационных технологий. 2025. Т. 6. No. 22. Стр. 37–56. (На англ.). https://doi.org/10.54309/IJICT.2025.22.2.003.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов

Introduction

The modern construction industry is one of the largest and most strategically significant sectors of the economy, driving infrastructure development and providing employment for a substantial portion of the population. However, the level of automation and adoption of advanced technologies in construction remains significantly lower compared to industries such as manufacturing, healthcare, or transportation (Kozlovska et al., 2021: 4–10; Shalbolova et al., 2023: 3–5) research indicates that the construction sector is characterized by low productivity, a high number of incomplete or delayed projects, and significant safety risks for workers. According to KPMG's 2023 report, up to 70 % of construction companies face considerable challenges in real-time monitoring of key construction metrics. This results in resource overruns and schedule deviations.

Traditional methods for monitoring construction processes, quality, and safety rely on visual inspections, manual data entry, and report analysis, which introduce subjectivity, documentation errors, and increased time costs. Although video surveillance systems are increasingly being deployed on construction sites, they primarily operate in a passive mode and require continuous human intervention for video stream analysis, limiting their overall effectiveness. Automated construction monitoring systems powered by artificial intelligence (AI) are still in the early stages of development. However, there are already successful cases of their practical implementation, demonstrating significant potential for improving safety and efficiency in construction operations (Vukicevic et al., 2024: 2–6).

However, despite its significant potential, existing construction monitoring methods still face several unresolved challenges. Most AI architectures require large volumes of labeled data for effective training, yet specialized datasets for construction sites (e.g., the Construction Site Safety Computer Vision Project) remain limited in terms of the number of images and scene diversity. Additionally, there is no unified system for integrating various monitoring tools. Construction companies often implement fragmented video analysis solutions, while comprehensive platforms that integrate data from drones, IP cameras, trail cameras, and sensor systems into a single intelligent decision-support environment have yet to be developed at a sufficient level. Moreover, existing algorithms lack adequate adaptation to the complex conditions of construction sites. The high visual dynamics of construction scenes, adverse weather conditions, poor lighting, and various obstacles (e.g., scaffolding and trees) pose significant challenges for current object detection and segmentation algorithms (Tahir et al., 2024: 103; Kumar et al., 2023: 20–23). Another critical aspect is real-time processing of video streams and images, as this requires substantial computational resources, making it impractical to deploy complex AI models without leveraging cloud-based GPU solutions.

In this context, there is a growing need for the development of an intelligent construction monitoring system that integrates machine learning methods, cloud technologies, and video image analysis.

The objective of this study is to design an automated system capable of real-time data processing from IP cameras, trail cameras, and drones to monitor construction safety, track progress, and assess resource utilization efficiency.

To achieve this goal, the following tasks were undertaken:

1. Integration of data from various monitoring sources.
2. Development of a computer vision model for detecting personal protective equipment (PPE) and monitoring construction progress.
3. Creation of an intelligent platform for analyzing and visualizing incoming data.
4. Evaluation of the system's effectiveness under test deployment conditions.

To substantiate the relevance of the proposed approach, we will examine the current technological solutions available today and analyze their limitations in detail.

Modern digital technologies in construction are evolving rapidly; however, they remain fragmented and primarily address specific monitoring and management tasks. The most widely used tools today include Building Information Modeling

(BIM) platforms, cloud-based project management systems, video analytics modules, and IoT systems. For instance, Building Information Modeling (BIM) software such as Autodesk Revit, Tekla, and Bentley is effectively utilized for designing and planning construction projects. However, BIM systems do not provide automated real-time monitoring of construction progress and require continuous manual updates to maintain accurate project information.

Cloud-based project management platforms such as Procore, PlanGrid, and Aconex facilitate document management and administrative oversight but lack the ability to automatically monitor construction processes or analyze video data in real time. Their dependence on manual data entry significantly slows down the detection of violations and the decision-making process.

Computer vision and deep learning technologies are increasingly being utilized to automate construction site monitoring, including the detection of safety violations (e.g., absence of helmets and protective eyewear), construction progress tracking, and equipment condition inspection. Various neural network models, such as YOLO, U-Net, MobileNet, and Vision Transformers, are employed for these tasks, each offering different levels of accuracy and performance. However, their effective deployment in construction environments is constrained by a lack of specialized labeled datasets, the complexity of scenes (e.g., low lighting, object occlusions, dynamic backgrounds), and high computational requirements.

Moreover, existing cloud solutions such as AWS, Google Cloud Platform, and Azure are not yet fully optimized for real-time video stream processing from construction sites or come with high operational costs, making the widespread adoption of automated monitoring systems in the industry challenging.

Thus, despite advancements in individual technologies, the construction industry still faces several significant limitations: data fragmentation, lack of integrated platforms, absence of real-time automated analysis, and high computational costs. Overcoming these challenges requires an intelligent system capable of integrating multiple data sources (IP cameras, trail cameras, drones, and sensors) into a unified cloud platform that enables automated safety violation detection, construction progress monitoring, and effective adaptation to real-world construction site conditions.

In response to these challenges, there is a growing need for the development of an intelligent construction monitoring system that integrates machine learning methods, cloud technologies, and automated video analysis.

Materials and Methods

Data Sources and Equipment

Effective construction site monitoring requires a comprehensive approach to data collection, incorporating both stationary and mobile surveillance systems. Stationary IP cameras serve as a key component of the proposed system. Cameras are selected based on their support for the Real-Time Streaming Protocol (RTSP) and resolution sufficient for reliable detection and recognition of small objects. Following an extensive analysis, the Hikvision DS-2CD2043G2-I camera demonstrated the highest efficiency, as shown in Figure 1. This choice is justified by its high image quality, strong adaptation to challenging lighting conditions, and built-in intelligent features such as motion detection and real-time analytics. For optimal coverage, cameras should be strategically installed at critical points on the construction site, ensuring a wide field of view over key work areas and movement pathways for both personnel and machinery. Mounting at a height of no less than 3 meters minimizes interference from construction equipment and structures, enhancing detection accuracy.

Fig.1. Hikvision DS-2CD2043G2-I Camera (image sourced from the official Hikvision website)

An alternative approach to construction site monitoring involves the use of trail cameras, which enable autonomous surveillance in hard-to-reach areas, triggering upon motion sensor activation. During testing, the H881 and T31ZX 4G LTE trail cameras were evaluated, as shown in Figure 2. The H881 model was preferred due to its high image quality in dusty and low-light conditions, as well as its extended battery life. This model supports data storage on SD cards up to 128 GB, ensuring reliable offline operation. For situations where physical access to the camera is difficult, the T31ZX 4G LTE is recommended, as it enables image transmission via 4G/3G mobile networks. With an average resolution of 12–16 MP, each frame occupies approximately 3–5 MB, allowing for the storage of up to 30,000 images. Since real-time video streaming is not required and potential data transmission delays are acceptable, these cameras provide an effective solution for periodic construction monitoring.

To collect aerial data on construction sites, unmanned aerial vehicles (UAVs), commonly known as drones, are utilized. These devices provide wide-format photo and video capture and enable regular inspection of hard-to-reach areas. The use of drones facilitates construction progress monitoring, early detection of potential violations, and the creation of digital 3D models of structures. Following a comparative analysis of various drone models, the DJI Mini 3 Pro was selected, as shown in Figure 3. This decision was based on the mobility of the model (it is easily transportable, deploys quickly, and requires minimal flight preparation) and its 4K camera with a wide dynamic range, which ensures detailed imaging of construction sites. Additionally, the built-in SDK provides future opportunities for automating flight routes and implementing autonomous inspection missions.

a. Trail Camera H881

b. Trail Camera T31ZX c 4G LTE

Fig. 2. Installed Trail Cameras

a. Appearance

б. Drone Video Recording Process

Fig.3. Drone DJI Mini 3 Pro RC

System Architecture

Video and photographic materials from cameras and drones generate a substantial volume of data, necessitating the use of a cloud-based platform for storage, processing, and subsequent analysis. To address this, the system integrates with the Lightning AI cloud platform. Video footage from drones and IP cameras is automatically uploaded to the platform, where preprocessing operations are performed, including compression, normalization, and segmentation of video sequences into individual frames. The processed data is then utilized for training neural network models and real-time object recognition. This approach enables the rapid detection of safety violations on construction sites while ensuring efficient storage and analysis of large datasets without compromising system performance.

As part of this study, an integrated video monitoring platform for construction processes has been developed, aimed at automating control, analyzing construction dynamics, and ensuring safety. Figure 5 presents the high-level architectural concept of the system, which facilitates the acquisition, transmission, storage, and analysis of video data through the application of computer vision techniques and cloud computing. The proposed architecture is modular and scalable, allowing for continuous upgrades and the incorporation of new components as data accumulates, algorithms are optimized, and additional functionalities are integrated.

The main stages of data processing within the system include:

- Acquisition of images and video streams from IP cameras and other sources.
- Transmission of data to the server using the RTCP protocol, ensuring a reliable connection between devices and the cloud platform.
- Initial image processing utilizing machine learning methods on a GPU-powered server.
- Storage of processed data and analysis results in a cloud-based object repository and associated database.
- Visualization of results through a web interface.

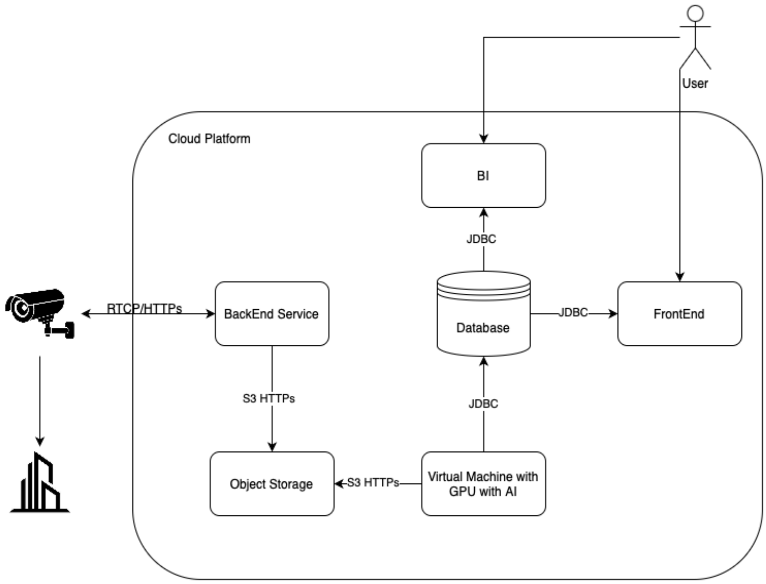

Fig. 5. Platform Architecture

The central component of the architecture is the backend service, which receives incoming video streams and images from all monitoring devices. It performs data routing, initial processing (such as noise removal and cropping), and stores both raw and processed information in the cloud storage for further analysis.

The training and inference module is responsible for processing video data using machine learning techniques. Its primary functions include training models based on incoming data, real-time object detection and classification, and periodic model updates to account for newly acquired data. To ensure efficient operation, the module runs on a cloud server equipped with a GPU. The training process is organized as a pipeline (CI/CD workflow) that includes stages such as loading new data from storage, data augmentation and normalization, neural network training, validation on a holdout dataset, and deployment of the improved model to replace the previous one. To facilitate this process, tools such as MLflow, TensorFlow Extended (TFX), and Kubeflow are integrated, enabling containerization and automated model updates in response to changing conditions. This approach ensures flexibility and adaptability of the algorithms over long-term system operation.

PostgreSQL is used as the primary database management system (DBMS) due to its reliability, scalability, and support for analytical operations. The database records information about detected objects, construction progress, and identified safety violations. PostgreSQL supports JSONB fields for storing complex structured data, which simplifies handling image metadata and inference results. Additionally, the database interacts with BI systems via JDBC connectors, enabling the generation of analytical reports and visualization of monitoring results.

For end users, dedicated interfaces have been developed to provide convenient access to monitoring results and analytics. The system consists of two main components:

- A BI dashboard that facilitates the visualization of key metrics, including safety violations, construction progress, and analysis of construction equipment utilization.
- A web application that provides real-time access to data, including viewing images and video recordings from cameras, detailed reports on recorded events, configuration of analytics parameters, and data filtering options.

Approaches to Image Analysis and Object Recognition

One of the key tasks of computer vision in construction monitoring is the detection and classification of objects such as construction structures, workers, specialized equipment, and personal protective equipment (helmets, vests, gloves). Various neural network architectures optimized for accuracy and processing speed in real-world conditions are employed for this purpose.

Object Detection

A critical task of the construction site monitoring system is the detection and classification of safety elements, such as construction helmets, vests, workers, and specialized equipment. To address the detection challenge, various state-of-the-art deep learning architectures optimized for accuracy and speed were evaluated and tested. Lightweight models, such as MobileNet, offer energy efficiency and low computational costs. However, experiments revealed a significant decline in their accuracy when dealing with complex scenes featuring numerous overlapping objects and dynamic backgrounds. As an alternative, Vision Transformer (ViT) models were considered, leveraging the self-attention mechanism to provide deep and effective analysis of visually complex scenarios, including cases of partial object occlusion. At the same time, ViT models are characterized by high computational demands and hardware requirements, which significantly limit their applicability for real-time video stream analysis.

The analysis also considered modern and well-established neural network architectures such as YOLO (You Only Look Once) and Faster R-CNN. As demonstrated in, the YOLO architecture, due to its use of multi-scale processing and efficient one-step detection, achieves high speed and good accuracy even under challenging shooting conditions, such as variable lighting and dynamic scenes. Meanwhile, the Faster R-CNN model (Shaoqing et al., 2017: 1137–1149), while providing high accuracy through its two-stage approach (region proposal generation followed by classification), exhibits significantly lower processing speed, making it less suitable for real-time monitoring tasks.

Based on the analysis, the YOLOv11 model was selected as the primary object detector in the system, as it provides high recognition accuracy with fast inference speed and supports efficient fine-tuning on specialized construction datasets.

Segmentation

For a detailed analysis of construction sites, the monitoring system employs image segmentation, enabling the identification of defect areas, tracking construction phases, and quantitatively assessing the progress of structural development. Various neural network architectures for segmentation were evaluated, including DeepLabV3+, Mask R-CNN, and U-Net. The study demonstrated that DeepLabV3+ models achieve high segmentation accuracy; however, their computational demands remain significant for real-time systems. Similarly, Faster R-CNN, extended with Mask R-CNN as described in, provides precise object detection but requires substantial computational resources, limiting its applicability in dynamic construction site environments.

In contrast to these models, U-Net (Olaf et al., 2015: 235–238) with its symmetrical encoder-decoder architecture, enables high-detail segmentation while maintaining relatively low computational costs. Practical studies indicate that applying U-Net to the analysis of construction site images can achieve segmentation accuracy of up to 90 % for defect areas, significantly reducing the subjectivity of manual inspection and facilitating the automation of construction progress assessment.

Based on this analysis, the U-Net architecture was selected for the segmentation module in the system, as it provides an optimal balance between accuracy and processing speed for analyzing construction-related images.

Classification

In the development of the monitoring system, special attention is given to the automatic detection of safety violations. Various approaches can be employed for this task, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrid models. Testing has confirmed that CNNs demonstrate high efficiency in pattern recognition and object classification in images. They can effectively extract spatial features of objects and exhibit strong robustness to various image distortions. Architectures such as ResNet (Kaiming et al., 2016: 770–778) and EfficientNet (Mingxing et al., 2019: 6105–6114) have relatively low model complexity, enabling real-time performance even under challenging imaging conditions.

The developed system employs the ResNet-50 model as the core classifier for safety violations. This model has demonstrated high accuracy (approximately 92 %) in classifying such situations on test data.

Data Preparation and Model Training

For the creation of training datasets, image annotation tools such as LabelImg and VGG Image Annotator (VIA) were utilized. These tools allowed for the manual labeling of object boundaries (e.g., workers, helmets, vests, machinery, etc.) in each image. Special attention was given to maintaining consistency in the annotation approach and accurately defining object boundaries, which is critical for the proper training of models.

At the initial stage of research, the primary sources of information included online resources and publicly available datasets depicting construction processes and workers on construction sites. These datasets contained both positive and negative examples for each category of violation (Table 1).

Table 1.1 Categories of Violations on Construction Sites

Dataset	Total Number of Images	Training Set	Validation Set	Test Set
Dataset 1 (PPE detection Dataset [25])	1126	788	226	112
Dataset 2 (PPE TEST Dataset [26]) 1624		1129	332	163
Dataset 3 (Personal Protective Equipment Dataset [27])	2092	1460	421	221
Total	4842	3371	979	496

Additionally, existing open datasets such as COCO (Tsung et al., 2014: 740–755) and PASCAL VOC were examined. For instance, the COCO-2017 dataset contains approximately 2,600 images annotated with the label "Person," as shown in Figure 6. However, most of these images are not related to the construction domain. Moreover, a dataset titled "Construction Workers," consisting of 684 annotated images specifically focused on construction workers, was obtained from the Roboflow Universe platform (Figure 7).

The compiled dataset includes photographs of workers with clearly visible personal protective equipment and attire, group images containing multiple workers to train the model in distinguishing individuals within a group, as well as images captured under various lighting and weather conditions. This data diversity enhances the model's robustness across different scenarios.

Fig. 6. Example of an Annotated Image from the COCO-2017 dataset

Fig. 7. Example of an Annotated Image from Roboflow Universe

An analysis of the final dataset revealed a significant class imbalance (Figure 8). The histogram illustrates that the "Hardhat" class is predominant, with over 30,000 annotated helmet instances, whereas categories like "Mask" or "NO-Goggles" contain significantly fewer examples. Such an imbalance in the dataset may lead to reduced model accuracy for underrepresented categories. Therefore, future research aims to increase the number of objects in these underrepresented classes.

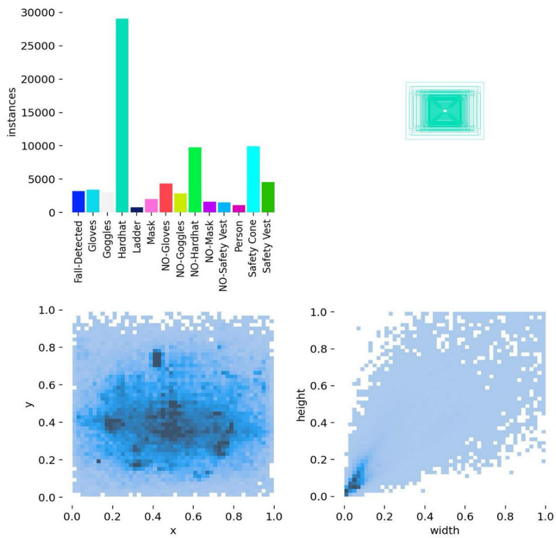

Fig. 8. Dataset Distribution by Classes

The heatmaps (Figure 9) illustrate the distribution of object center coordinates (x, y) and their normalized sizes (width, height). It can be observed that objects tend to appear closer to the center of the image and generally have relatively small width and height (up to 20–30% of the frame dimensions).

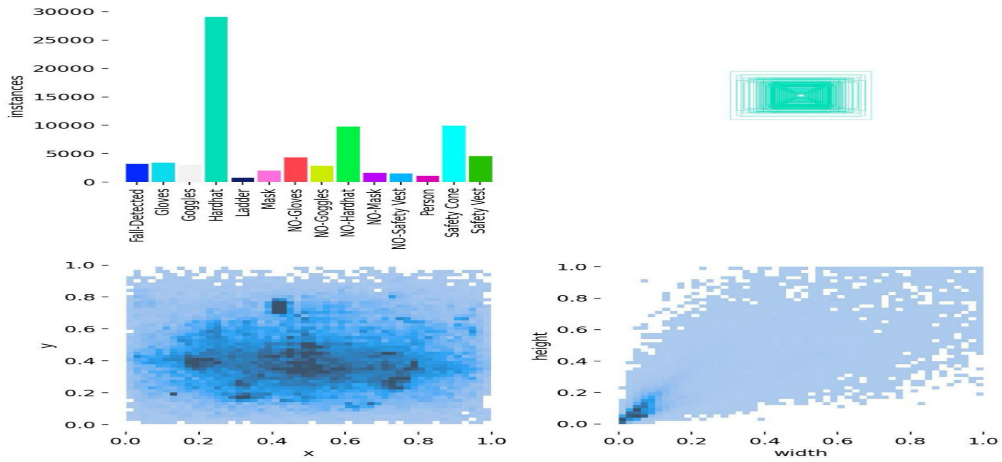

Fig.9. Distribution of Object Center Coordinates (x, y) and Their Normalized Sizes

For the recognition of personal protective equipment (PPE), the YOLOv11 model was utilized. This model inherits general features from pre-trained weights (ImageNet) and is further fine-tuned on a specialized dataset. The training process included image preprocessing (helmets, vests, goggles, etc.) with subsequent bounding box annotation, normalization and scaling to a resolution of 640×640, selection of optimal hyperparameters, training, and validation on the test set.

To ensure sufficient convergence while maintaining generalization ability, the model was trained for 50 epochs with a batch size of 16. The Adam optimizer was employed with an initial learning rate of 0.01 and a momentum coefficient of 0.937. This optimizer was chosen due to its fast convergence and stability on sparse features, which is particularly important for detecting small PPE objects. To mitigate overfitting, L2 regularization (weight_decay coefficient = 0.0005) was applied to constrain model weight growth. These parameters enabled stable model convergence without significant overfitting.

Results

To assess the effectiveness of the developed system, tests were conducted on a test dataset containing images of a construction site with annotated objects. For the trained YOLOv11 model (Figure 10), the mean Average Precision (mAP@0.5) was approximately 70 %. Specifically, for the “Person” class, an AP@0.5 of 75.1 % was achieved, while “Helmet” reached ~68 %, “Vest” ~72 %, and smaller objects such as “Phone” and “Cigarette” had an accuracy between 50–60 %. Under a stricter evaluation criterion of mAP@0.5:0.95, the average precision was ~35 %, considering different IoU thresholds. Additional calculations showed that for the “Person” class, Precision = 76.5 % and Recall = 71.6 %, resulting in an F1-score of approximately 74 %. The safety equipment classes also demonstrated precision and recall values exceeding 70 %.

Thus, the developed model produces minimal false positives; however, misclassification remains possible in more complex cases. That is, there are rare instances where non-human objects are incorrectly labeled as people, but there were also cases where real workers were missed in images with significant occlusions or poor image quality (Figure 11).

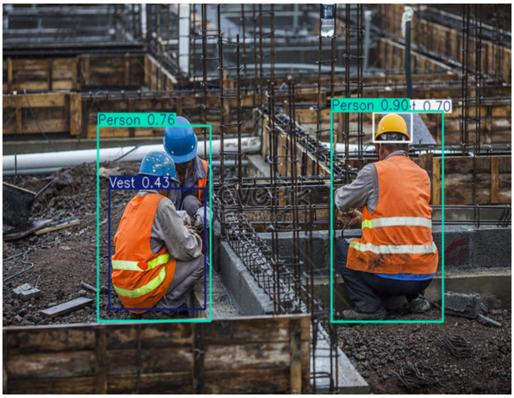

*Fig.10. E*xample of Worker Detection Using YOLOv11

Fig.11. Example of Missed Objects in the «Person» Class

For a deeper understanding of the model's performance, the obtained metrics were analyzed, and classifier characteristic graphs were constructed. Figures 12 and 13 present two versions of the confusion matrix (normalized and absolute), reflecting the distribution of correct and incorrect classifications for each predicted class. In both cases, the horizontal axis represents the true class labels, while the vertical axis shows the labels predicted by the model. The highest classification results on the main diagonal (0.90–0.98 percent) were achieved for the "Hardhat" and "Goggles" classes. The "Mask" class demonstrates significantly lower performance (around 0.35), which may be due to visual obstructions or face occlusions. The absolute error values (Figure 13) show that the most common type of error is misclassification between the presence and absence of a hardhat (over 300 errors between the "Hardhat" and "No-Hardhat" classes). This data clearly indicates the need to improve the model for rare classes and fine-tune the detection thresholds. Analyzing the confusion matrix helps identify which categories the algorithm is most reliable for (e.g., hardhats and goggles) and which still have a high error rate (e.g., masks and missing hardhats). Such information is crucial for further dataset balancing and optimizing the model's sensitivity for different object types.

In Figure 14, the graph shows the dependence of the F1 score on the detector's Confidence threshold. It is evident that as the threshold increases up to a certain value, the F1 score improves, but beyond that point, it starts to decline. The maximum F1 (~0.73) is achieved at a confidence threshold of approximately 0.27. This result indicates that the optimal trade-off between recall and precision for this model is reached at a confidence level of around 0.25–0.30.

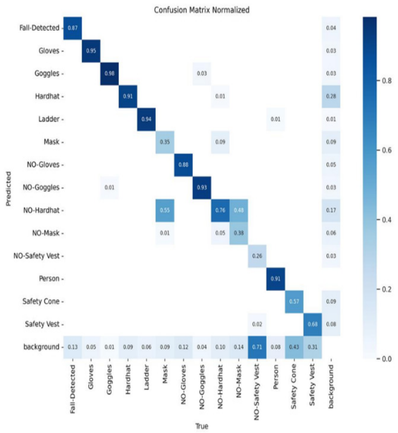

Fig. 12. Normalized Confusion Matrix of Classification Results

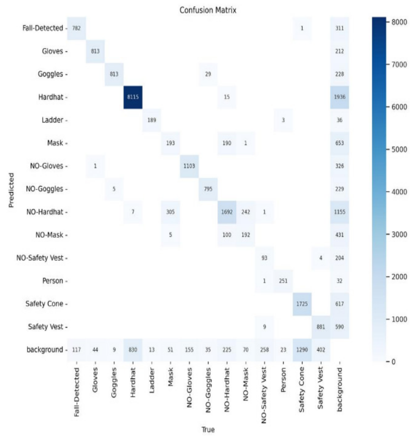

*Fig.13.*Confusion Matrix with the Absolute Number of Predictions

At lower thresholds, the model detects more objects (higher recall), but false positives increase, reducing precision. Conversely, at higher thresholds, precision improves at the cost of missing some objects. Thus, the chosen range of ~0.27 for the confidence threshold is the most effective, confirming the necessity of proper detector calibration.

In Figure 15, the Precision-Recall curves for each class, as well as for the combined case ("all classes"), are shown. The highest scores (0.951 and 0.949) are observed for the "Ladder" and "Goggles" classes, respectively, while the "Mask" class achieves only 0.467, indicating challenges in detecting masks. At IoU=0.5, the model reaches an average detection accuracy of 75.4% across all classes. These results demonstrate that the model performs well for most categories; however, classes with fewer training examples (e.g., masks) lag significantly in quality.

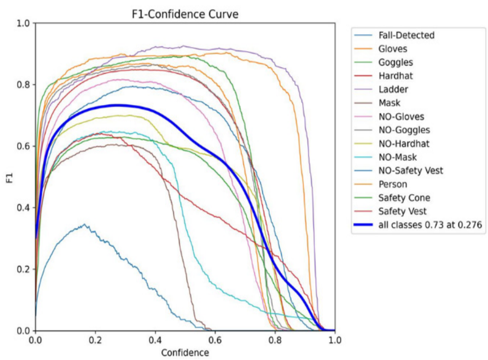

Fig.14. F1-Confidence Curve

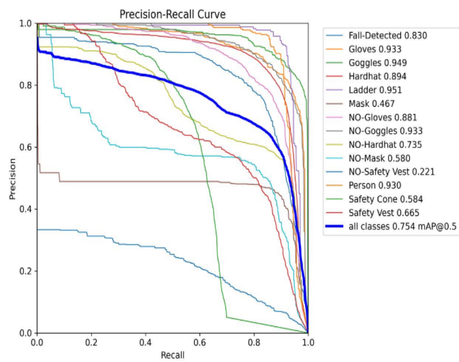

*Fig.15.*Precision-Recall Curve

To evaluate the training process of the YOLO model, curves showing the changes in loss functions and key metrics on the training and validation datasets were plotted, as shown in Figure 16:

- train/box_loss - change in the loss function on the training set for bounding box coordinate regression.
- train/cls_loss represents the training classification
- train/dfl_loss – The loss value associated with distribution focal loss, which is used to refine coordinate regression in YOLO).
- metrics/precision(B) – the accuracy metric on the training dataset (or in “best” mode)- metrics/recall(B) it represents recall, i.e., the proportion of objects detected by the model among all actual objects.
- The metric val/box_loss is analogous to train/box_loss, but it is calculated on the validation set.
- val/cls_loss - classification loss on the validation set.
- val/dfl_loss - validation part of the distribution focal loss.
- metrics/mAP50(B) - mean Average Precision at an IoU threshold of 0.5
- metrics/mAP50-95(B) - a stricter metric averaging mAP across IoU thresholds from 0.5 to 0.95.

Overall, these curves demonstrate stable training convergence: all types of losses (box, cls, dfl) consistently decrease, while key metrics (Precision, Recall, mAP) increase. Moreover, the training and validation curves follow each other quite closely. This indicates the absence of clear signs of overfitting and suggests good generalization performance of the model on data that was not involved in training.

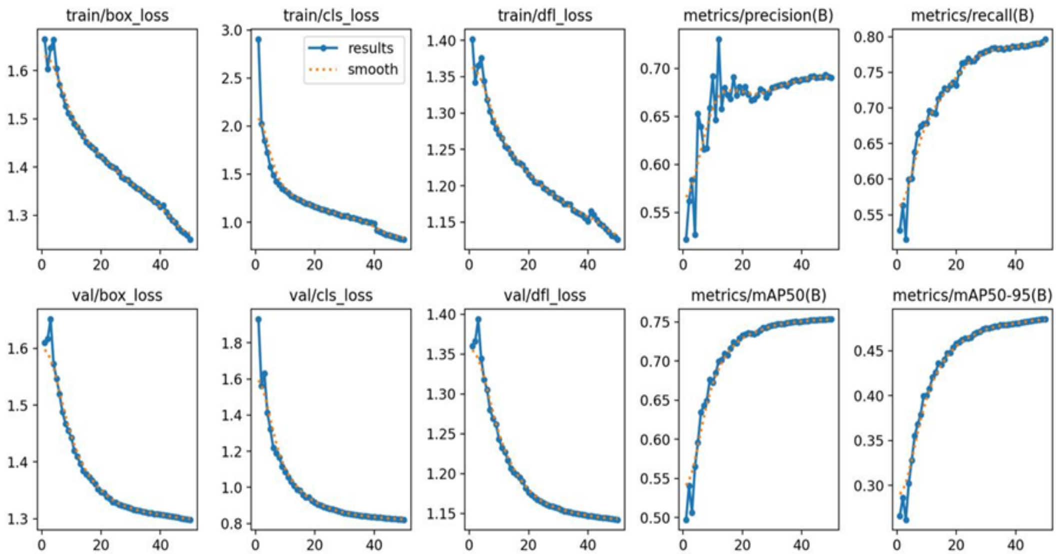

Fig.16. Loss and Key Metric Curves (Train/Val Loss, Precision, Recall, mAP

Discussion of the Results

The developed system successfully addresses the tasks of automatic detection and classification of key objects (workers and personal protective equipment), enabling (almost) real-time detection of potential safety violations. High Precision (0.90 and above) and Recall (0.80–0.85) for the classes "Hardhat" and "Goggles" indicate the reliability of the detector when sufficient training examples and favorable shooting conditions are provided.

The analysis of the confusion matrix results (see Figures 12 and 13) confirms that the model demonstrates high accuracy for frequently occurring classes in the dataset. However, objects such as "Mask" or "NO-Hardhat" are detected less reliably. This points to an issue of sample imbalance and the difficulty of recognizing occluded objects. Additionally, further experiments showed that the optimal Confidence threshold for maximizing the F1-score is around 0.25–0.30, as confirmed by the curves in Figure 14. This threshold selection ensures the minimization of both missed objects and false positives. Information about the preferred threshold will be useful when deploying the system in practice, allowing for tuning the detection sensitivity to meet the specific requirements of a given site.

The dynamics of the training and validation loss metrics (Train/Val Loss) shown in Figure 16 indicate the stable convergence of the model. There is a smooth decrease in all types of losses (box_loss, cls_loss, dfl_loss) and a parallel increase in mAP. The validation curves closely follow the training curves, which suggests that there is no significant overfitting, and the model has an acceptable generalization ability. However, with stricter IoU criteria (e.g., mAP@0.75 and above), the model's accuracy decreases significantly, indicating the need for further refinement of bounding box coordinates and additional augmentations simulating complex shooting angles.

Despite the results achieved, this study has several limitations. The accuracy of recognition is heavily influenced by the shooting conditions: lighting, angle, weather factors, and the presence of noisy backgrounds. The existing class imbalance leads to a decrease in detection recall for rarely occurring categories. Additionally, it should be noted that the system is intended for static or slightly moving scenes; when using sharp camera movements (e.g., when mounted on a drone), special attention should be paid to stabilization algorithms and adaptive data augmentation.

To further improve recognition quality, it is planned to increase the proportion of rare categories in the training dataset, utilize more advanced augmentation techniques including modeling complex lighting, noisy backgrounds, and occluded objects. Research on Vision Transformers and hybrid architectures is also planned to potentially improve accuracy in complex and dynamic scenes. To optimize computational costs, it is necessary to study the adaptation of input image size and detection thresholds to specific requirements (real-time operation or periodic sampling).

Conclusion

As a result of the conducted research, a comprehensive monitoring system architecture was designed, based on the integration of photo and video data acquisition devices with the Lightning AI platform, optimized for deep learning tasks. The proposed solution demonstrated high efficiency in test conditions at construction sites: when using YOLO and ResNet-50 models, the detection accuracy of workers and personal protective equipment reached 70–75 % (mAP@0.5), and the identification of violations exceeded 90 %. The integration of machine learning methods and cloud services ensures scalability and high performance in video stream analysis, but it requires consideration of network bandwidth and adherence to strict data protection requirements.

The further development of the system involves expanding the datasets to achieve a balanced representation of different object classes, as well as including new scenarios related to additional sensors (e.g., seismic) and complex shooting conditions. A promising direction is the study of modern computer vision architectures capable of improving detection and segmentation accuracy in real-time. At the same time, network infrastructure characteristics need to be considered, as working with multiple Full HD cameras simultaneously can generate data streams of tens of megabits per second. Therefore, it is important to use reliable wired Gigabit Ethernet connections or high-speed wireless solutions. As the system scales, special attention should be paid to information security and privacy: regular data backups and distributed storage must be implemented, as well as strict encryption and access control policies.

Thus, building an intelligent monitoring platform for construction sites requires a comprehensive approach: improving computer vision algorithms, integrating additional data sources, advancing network technologies, and ensuring cybersecurity. This approach will significantly enhance the efficiency and safety of construction processes, opening new opportunities for the digital transformation of the industry.

REFERENCES

Hasirlioglu S., Riener A. (2020). Challenges in Object Detection Under Rainy Weather Conditions: Second EAI International Conference, INTSYS. — Guimarães, Portugal. Proceedings. From https://10.1007/978-3-030-14757-0_5. (in Eng.).

Josh T., Rachel F., Alex R., Jonas S., Wojciech Z., Pieter A. (2017). Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World, from https://10.48550/arXiv.1703.06907. (in Eng.).

Kozlovska M., Klosova D., Strukova Z. (2021). Article: Impact of Industry 4.0 Platform on the Formation of Construction 4.0 Concept: A Literature Review. Sustainability. — Basel, Switzerland. — Pp. 4–10. DOI: https://doi.org/10.3390/su13052683 (in Eng.).

Kaiming H., Georgia G., Piotr D., Ross G. (2017)"Mask R-CNN," IEEE International Conference on Computer Vision (ICCV). — Venice, Italy. — Pp. 770–778. DOI: https://doi.org/10.1109/ICCV.2017.322. (in Eng.).

Kumar D., Muhammad N. (2023). Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8. — Basel, Switzerland. — Pp. 20–23. DOI: https://doi.org/10.3390/s23208471. (in Eng.).

KPMG global tech report. — Retrieved 10 February. — 2025, from: https://kpmg.com/xx/en/our-insights/ai-and-technology/kpmg-global-tech-report-2023.html (in Eng.).

Kaiming H., Xiangyu Zh., Shaoqing R., Jian S. (2016). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), from https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html (accessed: 10.02.2025) (in Eng.).

Mingxing T., Quoc V.L. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In Proceedings of the 36th International Conference on Machine Learning (ICML). — Pp. 6105–6114. DOI: https://doi.org/10.48550/arXiv.1905.11946 (in Eng.).

Olaf R., Philipp F., Thomas B. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention. — Springer, Cham. — Pp. 235–238. DOI: https://doi.org/10.1007/978-3-319-24574-4_28 (in Eng.).

Tahir N.U.A., Zhang Z., Asim M., Chen J., ELAffendi M.M. (2024). Object Detection in Autonomous Vehicles under Adverse Weather: A Review of Traditional and Deep Learning Approaches. Algorithms. — Basel, Switzerland. — Pp. 103. DOI: https://doi.org/10.3390/a17030103 (in Eng.).

Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, Piotr Dollár (2014). Microsoft COCO: Common Objects in Context. European Conference on Computer Vision (ECCV). — Pp. 740–755. from https://arxiv.org/pdf/1405.0312 (in Eng.).

Tsung-Yi L., Michael M., Serge B., Lubomir B., Ross G., James H., Pietro P., Deva R., Lawrence Zitnick C., Piotr D. (2014). Microsoft COCO: Common Objects in Context. European Conference on Computer Vision (ECCV). — Pp. 740–755. From https://arxiv.org/pdf/1405.0312 (in Eng.).

Shaoqing R., Kaiming H., Ross G., Jian S. (2017). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence. — Pp. 1137–1149. DOI: https://doi.org/10.1109/TPAMI.2016.2577031 (in Eng.).

Shalbolova U., Junussova A., Salykov A. (2023). Efficiency of Digitalization of Housing and Communal Services and Construction in Kazakhstan. — Pp. 3–5. DOI: https://doi.org/10.1007/978-981-99-4229-9_5. (in Eng.).

Vukicevic A.M., Petrovic M., Milosevic P. et al. (2024). A systematic review of computer vision-based personal protective equipment compliance in industry practice: advancements, challenges and future directions. Artif Intell Rev 57. — 319. — Pp. 2–6. DOI: https://doi.org/10.1007/s10462-024-10978-x (in Eng.).

Yuqing W., Zhaoliang X., Xinlong W., Chunhua S., Baoshan Ch., Hao S., Huaxia X. (2020). End-to-End Video Instance Segmentation with Transformers. https://doi.org/10.48550/arXiv.2011.14503. (in Eng.).

ХАЛЫҚАРАЛЫҚ АҚПАРАТТЫҚ ЖӘНЕ КОММУНИКАЦИЯЛЫҚ ТЕХНОЛОГИЯЛАР ЖУРНАЛЫ

МЕЖДУНАРОДНЫЙ ЖУРНАЛ ИНФОРМАЦИОННЫХ И КОММУНИКАЦИОННЫХ ТЕХНОЛОГИЙ

INTERNATIONAL JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGIES

Правила оформления статьи для публикации в журнале на сайте:
https://journal.iitu.edu.kz
ISSN 2708–2032 (print)
ISSN 2708–2040 (online)
Собственник: АО «Международный университет информационных технологий» (Казахстан, Алматы)

Подписано в печать 15.06.2025.
Формат 60х881/8. Бумага офсетная. Печать - ризограф. 9,0 п.л. Тираж 100
050040 г. Алматы, ул. Манаса 34/1, каб. 709, тел: +7 (727) 244-51-09).

Издание Международного университета информационных технологий
Издательский центр КБТУ, Алматы, ул. Толе би, 59