

UDC 004.896
CSCSTI 28.23.29

DOI: [10.54859/kjogi109018](https://doi.org/10.54859/kjogi109018)

Received: 22.06.2026.

Accepted: 30.07.2026.

Published: 30.09.2026.

Original article

Digital Monitoring of The Use of Personal Protective Equipment at Industrial Facilities Using Neural Network Architectures

Galymzhan S. Abdimanap^{1,2}, Anel N. Alimova¹, Kairat A. Bostanbekov¹,
Renat Ye. Abdrakhmanov¹, Dana Orenkyzy¹, Daniyar B. Nurseitov^{1,2}

¹KMG Engineering, Astana, Kazakhstan

²Satbayev University, Almaty, Kazakhstan

ABSTRACT

Background: Ensuring employees occupational safety at the oil and gas industry facilities remains a relevant objective, as far as traditional monitoring methods for the use of Personal Protective Equipment (PPE) are based on manual visual inspections and are susceptible to human factor. Most existing computer vision systems are limited to detecting only a small number of 2–6 classes of PPE categories and to verifying the anatomical consistency between detected protective equipment and the corresponding body parts of employees.

Aim: Developing and validating a digital monitoring method for compliance with PPE use requirements based on neural network architectures, integrating algorithms for object detection, human pose estimation, and anatomical matching of PPE elements in real time mode.

Materials and methods: A unique dataset of 16352 images (after augmentation) containing 13 object classes, including 6 types of PPE and 6 negative classes was created for training the model. The YOLOv8 model was used for object detection, and HRNet for human pose estimation. A two-tier video stream processing architecture was implemented, combining object tracking (BoT-SORT), spatial and anatomical matching and TensorRT quantization to improve system performance.

Results: During the training phase, the YOLOv8 model achieved Precision = 0.98, Recall = 0.97, and F1 = 0.94. When testing the developed system on 16 video files obtained from industrial sites, the system achieved a precision of 96.58%, a recall of 68.48%, and an F1 score of 0.8014. Detecting small objects (gloves) in cropped images improves detection efficiency by 2-3 times compared to full-frame processing.

Conclusion: The developed approach, combining the YOLOv8 and HRNet models and an anatomical matching algorithm, provides effective digital monitoring of compliance with PPE requirements in real-world production conditions. The obtained results confirm the potential of its application in the development of intelligent industrial safety monitoring systems at industrial facilities.

Keywords: big industrial safety; personal protective equipment; computer vision; YOLOv8 neural network; HRNet; human pose estimation; object detection.

To cite this article:

Abdimanap GS, Alimova AN, Bostanbekov KA, et al. Digital Monitoring of The Use of Personal Protective Equipment at Industrial Facilities Using Neural Network Architectures. *Kazakhstan journal for oil & gas industry*. 2026;8(3):76–89. DOI: [10.54859/kjogi109018](https://doi.org/10.54859/kjogi109018).

УДК 004.896
МРНТИ 28.23.29

DOI: [10.54859/kjogi109018](https://doi.org/10.54859/kjogi109018)

Получена: 22.06.2026.

Одобрена: 30.07.2026.

Опубликована: 30.09.2026.

Оригинальное исследование

Цифровой мониторинг использования средств индивидуальной защиты на промышленных объектах на основе применения нейросетевых архитектур

Ғ.С. Әбдіманап^{1,2}, А.Н. Алимова¹, К.А. Бостанбеков¹, Р.Е. Абдрахманов¹,
Д. Өренқызы¹, Д.Б. Нурсейтов^{1,2}

¹КМГ Инжиниринг, г. Астана, Казахстан

²КазННТУ им. К.И. Сатпаева, г. Алматы, Казахстан

АННОТАЦИЯ

Обоснование. Обеспечение производственной безопасности работников на объектах нефтегазовой отрасли является актуальной задачей, поскольку традиционные методы контроля использования средств индивидуальной защиты (далее – СИЗ) основаны на визуальных осмотрах и подвержены влиянию человеческого фактора. Существующие системы компьютерного зрения, как правило, ограничиваются детекцией 2–6 классов СИЗ и не учитывают анатомическое соответствие защитного оборудования частям тела работника.

Цель. Разработка и апробация метода цифрового мониторинга соблюдения требований по использованию СИЗ на основе нейросетевых архитектур, интегрирующего алгоритмы обнаружения объектов, оценки позы человека и анатомического матчинга элементов СИЗ в режиме реального времени.

Материалы и методы. Для обучения модели создан собственный уникальный датасет из 16352 изображений (после аугментации), содержащий 13 классов объектов, включая 6 типов СИЗ и 6 негативных классов. Для детекции объектов использована модель YOLOv8, для оценки позы человека – HRNet. Реализована двухуровневая архитектура обработки видеопотока, объединяющая трекинг объектов (BoT-SORT), пространственный и анатомический матчинг, а также квантизацию TensorRT для повышения производительности системы.

Результаты. На этапе обучения модель YOLOv8 достигла значений Precision = 0,98, Recall = 0,97 и F1 = 0,94. При тестировании разработанной системы на 16 видеофайлах, полученных с месторождений, достигнуты Precision = 96,58%, Recall = 68,48% и F1-меры = 0,8014. Детекция мелких объектов (перчатки) на обрезанных изображениях позволяет повысить эффективность обнаружения объектов в 2–3 раза по сравнению с полнокадровой обработкой.

Заключение. Разработанный подход, объединяющий модели YOLOv8, HRNet и алгоритм анатомического матчинга, обеспечивает эффективный цифровой мониторинг соблюдения требований по использованию СИЗ в реальных производственных условиях. Полученные результаты подтверждают перспективность его применения при создании интеллектуальных систем мониторинга производственной безопасности на промышленных предприятиях.

Ключевые слова: производственная безопасность, средства индивидуальной защиты, компьютерное зрение, нейросеть YOLOv8, HRNet, оценка позы человека, детекция объектов.

Как цитировать:

Әбдіманап Ғ.С., Алимова А.Н., Бостанбеков К.А., и др. Цифровой мониторинг использования средств индивидуальной защиты на промышленных объектах на основе применения нейросетевых архитектур // Вестник нефтегазовой отрасли Казахстана. 2026. Том 8, №3. С. 76–89. DOI: [10.54859/kjogi109018](https://doi.org/10.54859/kjogi109018).

ӨОЖ 004.896

FTAXP 28.23.29

DOI: [10.54859/kjogi109018](https://doi.org/10.54859/kjogi109018)

Қабылданды: 22.06.2026.

Мақұлданды: 30.07.2026.

Жарияланды: 30.09.2026.

Түпнұсқа зерттеу

Нейрондық желі архитектураларын қолдану негізінде өнеркәсіптік объектілерде жеке қорғаныс құралдарын пайдалануды сандық мониторингтеу

Ғ.С. Әбдіманап^{1,2}, Ә.Н. Әлімова¹, Қ.А. Бостанбеков¹, Р.Е. Әбдірахманов¹,
Д. Өренқызы¹, Д.Б. Нұрсейітов^{1,2}

¹ҚМГ Инжиниринг, Астана қаласы, Қазақстан

²Қ.И. Сәтбаев атындағы ҚазҰТЗУ, Алматы қаласы, Қазақстан

АНДАТПА

Негіздеу. Мұнай-газ саласы объектілеріндегі қызметкерлердің өндірістік қауіпсіздігін қамтамасыз ету өзекті міндет болып табылады, себебі жеке қорғаныс құралдарын (ЖҚҚ) пайдалануды бақылаудың дәстүрлі әдістері визуалды тексерулерге негізделген және адами фактордың әсеріне ұшырайды. Қолданыстағы компьютерлік көру жүйелері, әдетте, ЖҚҚ-ның 2–6 класын ғана анықтаумен шектеледі және қорғаныс құралдарының қызметкер дене бөліктеріне анатомиялық сәйкестігін ескермейді.

Мақсаты. Нейрондық желі архитектуралары негізінде ЖҚҚ пайдалану талаптарының сақталуын сандық мониторингтеу әдісін әзірлеу және апробациялау, нақты уақыт режимінде объектілерді анықтау, адамның позасын бағалау және ЖҚҚ элементтерінің анатомиялық сәйкестендіру алгоритмдерін біріктіру.

Материалдар мен әдістер. Модельді оқыту үшін 16352 кескіннен тұратын өзіндік бірегей деректер жиынтығы (аугментациядан кейін) құрылды, ол ЖҚҚ-ның 6 түрін және 6 теріс класты қамтитын 13 объект класынан тұрады. Объектілерді детекциялау үшін YOLOv8 моделі, ал адамның позасын бағалау үшін HRNet пайдаланылды. Жүйенің өнімділігін арттыру мақсатында объектілерді бақылауды (BoT-SORT), кеңістіктік және анатомиялық сәйкестендіруді, сондай-ақ TensorRT квантизациясын біріктіретін бейнеағынды өңдеудің екі деңгейлі архитектурасы іске асырылды.

Нәтижелері. Оқыту кезеңінде YOLOv8 моделі Precision = 0,98, Recall = 0,97 және F1 = 0,94 мәндеріне қол жеткізді. Кен орындарынан алынған 16 бейнефайлда әзірленген жүйені тестілеу барысында Precision = 96,58%, Recall = 68,48% және F1-өлшемі = 0,8014 нәтижелеріне қол жеткізілді. Кесілген кескіндердегі ұсақ объектілерді (қолғаптарды) детекциялау толық кадрлы өңдеумен салыстырғанда объектілерді анықтау тиімділігін 2–3 есе арттыруға мүмкіндік береді.

Қорытынды. YOLOv8 және HRNet модельдерін, сондай-ақ анатомиялық сәйкестендіру алгоритмін біріктіретін әзірленген тәсіл нақты өндірістік жағдайларда ЖҚҚ пайдалану талаптарының сақталуын тиімді сандық мониторингтеуді қамтамасыз етеді. Алынған нәтижелер оның өнеркәсіптік кәсіпорындарда өндірістік қауіпсіздікті мониторингтеудің зияткерлік жүйелерін құру кезінде қолданылу перспективаларын растайды..

Негізгі сөздер: өндірістік қауіпсіздік, жеке қорғаныс құралдары, компьютерлік көру, YOLOv8 нейрондық желісі, HRNet, адамның позасын бағалау, объектілерді детекциялау.

Дәйексөз келтіру үшін:

Әбдіманап Ғ.С., Әлімова Ә.Н., Бостанбеков Қ.А., және б. Нейрондық желі архитектураларын қолдану негізінде өнеркәсіптік объектілерде жеке қорғаныс құралдарын пайдалануды сандық мониторингтеу // Қазақстанның мұнай-газ саласының хабаршысы. 2026. 8 том, №3, 76–89 б.

DOI: [10.54859/kjogi109018](https://doi.org/10.54859/kjogi109018).

Introduction

In modern industry employee safety is a key element of effective process organization. Failure to comply with personal protective equipment (PPE) usage regulations poses a direct threat to the health and lives of employees, leading to the implementation of innovative technologies to automate the PPE use monitoring. Traditional approaches based on manual visual inspections and periodic checks are susceptible to human error and delays in identifying safety rules violation [1, 2]. Today, there are two main approaches to PPE monitoring such as sensor-based methods based on radio frequency identification and the Internet of Things [1, 2] and machine vision methods using deep learning algorithms [3–15]. The second group of methods has attracted particular attention due to their ability to analyze visual data without the involvement of additional means. However, existing solutions are primarily limited to detecting a few classes of objects and are not always capable of verifying the correct wearing of PPE.

The purpose of the study is to develop a digital monitoring method for compliance with PPE use standards at industrial sites, combining employee detection (YOLOv8), human pose estimation (HRNet), detection of 13 classes of PPE objects, and anatomical matching to verify the correct wearing of PPE in real time mode.

Literature review

This section reviews key research in the field of automated PPE detection using deep learning methods. The reviewed works reflect the main development directions in this field, including the application of deep learning algorithms, transfer learning methods and various approaches to improving PPE detection accuracy.

Xiong R. and Tang P. [4] proposed a method for detecting compliance with PPE use at construction yards based on employee pose analysis using pose estimation technology, instead of traditional object detection. In the proposed approach, the employee's body is represented as a skeletal model, and individual body parts are used as spatial landmarks to define part attention regions where the presence of relevant PPE elements (e.g., a hard hat on the head, a vest on the upper body) are expected to be worn. To localize these regions, the authors developed a set of rules based on human anatomical features, after which they used convolutional neural networks to classify the extracted image fragments.

Experimental results demonstrated the high efficiency of the proposed approach: the F1-score was 96.95% for detecting hard hats and 94.61% for detecting vests. Furthermore, the developed method demonstrated higher accuracy compared to traditional object detection models, including Faster R-CNN and YOLO. These results demonstrate the potential of using human pose estimation methods for automated compliance control

of industrial safety requirements and monitoring the use of PPE at construction yards.

A number of studies have been devoted to the application of traditional object detection methods for monitoring the use of PPE. In the work of Chen S. and Demachi K. [5], an approach to monitoring compliance with requirements for the use of PPE at nuclear power plant facilities was proposed. To train the model, an annotated dataset was formed, including 3808 images obtained both from webcams installed at an industrial facility and from open internet sources. A separate manually collected dataset was used for testing. In the framework of the study, two classes of objects were considered: a protective helmet and a full-face protective mask. To solve the detection problem, a one-stage YOLOv3 model with the basic Darknet-53 architecture was used. Training was performed in two stages: the first stage involved training with frozen convolutional layers of the basic network, and the second stage involved tweaking the model with unfreezing of all layers. To optimize the training process, the Adam optimizer algorithm with a batch size of 8 and different learning rates at individual stages was used. Based on experimental results, the developed model demonstrated an accuracy of 97.64% and a recall of 93.11%, processing a video stream in real time at 7.96 frames per second (FPS). However, the study is limited to monitoring only two types of PPE and does not consider a wider range of PPE. The authors note the need to expand the dataset and implement the developed model in real-world production conditions as future directions for development.

In a study by Delhi V.S.K., et al. [6], the problem of using PPE on construction yards in real time mode was examined using deep learning and computer vision methods to improve industrial safety. To train the model, the authors manually compiled a dataset consisting of approximately 2 500 images collected from open internet sources. The images were divided into four classes: "SAFE," "UNSAFE," "no clothing" and "no jacket."

The YOLOv3 architecture was used as the object detection model. To improve the robustness and generalization ability of the model, image augmentation methods were used during the data preparation stage, including horizontal flipping and rotation by up to 30°. The dataset was divided into training, validation, and test sets at a ratio of 90%, 8%, and 2%, respectively. Experimental results demonstrated the high efficiency of the proposed model then the mAP, Recall, and F1-score values reach 97%.

Wang Z., et al. [7] investigated the problem of employee compliance with occupational safety requirements using deep learning neural network methods adapted for real-time mode object detection. The authors developed and compared detection models based on the YOLOv3, YOLOv4, and YOLOv5 architectures for recognizing six classes of PPE objects. The CHV dataset [8] containing

1 330 high-quality images was used for training and testing the models. The dataset included six classes of objects: protective helmets of four colors, employee face images and vests. A comparative analysis revealed that the YOLOv5x model achieved the highest detection accuracy, reaching 86.55%, while the YOLOv5s model demonstrated the highest processing speed reaching up to 52 frames per second. The obtained results indicate a tradeoff between detection accuracy and computational efficiency.

Ma L., et al. [9] proposed a combined PPE detection algorithm using a portable version of the YOLOv4 model. A dataset containing approximately 25 000 images obtained from CCTV cameras installed at construction yards was generated to perform experiments. The images were unevenly distributed across six object classes and divided into training and test sets. Two object detection architectures were considered such as YOLOv4 and YOLOv4-Tiny. To improve the accuracy of the models, fine-tuning and optimization were performed using the original dataset. The best results were obtained using the CLSIm YOLOv4 model, based on the Channel and Layer Slimming (CLSIm) method. This approach contributed to reduce the number of model parameters by 98.2%, and to increase the output speed by 1.8 times, and, in the same time, to reduce the mAP value by only 2.1% compared to the baseline model. The results obtained demonstrate the high efficiency of the CLSIm method in reducing computational power and increasing the speed of object detection. In the study by Lo J.-H., Lin L.-K., and Hung C.-C. [10], the problem of developing deep learning algorithms for detecting the use of PPE in real time mode was considered. A dataset containing over 11 000 images was generated for training and evaluating the models. The capabilities of the YOLOv3, YOLOv4, and YOLOv7 architectures were investigated. To improve training quality and reduce the risk of overfitting, data preprocessing methods were used, including image flipping, cropping, noise addition, and color space transformation. A comparative analysis showed that the YOLOv7 model demonstrates the best results, achieving a mAP value of 97.29% with a processing speed (FPS) of 25.02 frames per second. The implemented system achieved a PPE detection accuracy of 97.5%, confirming the effectiveness of modern deep learning models for monitoring safety compliance at industrial facilities in real time mode. The authors note the need to further expand the dataset by including more PPE use scenarios, images with overlaying objects, and negative examples, as well as by improving data collection methods. It should be noted that most images in the dataset were captured under good lighting and favorable weather conditions, which may limit the generalization ability of the model and its effectiveness in more complex industrial conditions.

In the study by Lee Y.-R., et al. [11], a computer vision-based platform was proposed to monitor the proper use of PPE at construction yards. To compile the dataset, images with a resolution of 1280×720 pixels obtained from various sources, including Google Images, CCTV cameras and smartphones used at construction yards. The dataset included three classes of objects: "Person", "Hard Hat", and "Safety Vest". A total of 1 031 images were used for the model training, and 257 images were used for testing. The dataset was used to train a contemporary pixel-based PPE detection model of the YOLACT architecture. The DeepSORT algorithm was used to track objects between frames. The results of the study show that the YOLACT model achieved an mAP50 score of 66.4%, and the DeepSORT algorithm provided an accuracy of 91.3% in detecting the PPE use status. As a future research direction, the authors consider expanding the platform's functionality by identifying potentially hazardous situations on construction sites through analyzing the co-relation between employees, the PPE they wear, and the surrounding work environment.

Nugraha K.O.P.P. and Rifai A.P. [12] proposed an approach to monitoring compliance with PPE use requirements in industrial laboratories based on convolutional neural networks (CNN). The authors created their own dataset of images of laboratory employees containing various PPE wearing scenarios. A trained model classified the images with the aim of determination of the presence or absence of mandatory PPE items, including lab coats, safety glasses, and gloves. Experimental results showed that the developed approach provides acceptable accuracy in classifying elements under controlled operating conditions. However, the authors also found a significant dependence of classification quality on shooting conditions. In particular, changes in illumination and background characteristics leads to a noticeable decrease in recognition accuracy compared to the conditions presented in the training set. The obtained results demonstrate the existence of a domain shift problem inherent in PPE usage monitoring systems and highlight the need for more variable datasets covering a wider range of operating conditions, including different illumination levels, viewing angles, and environmental features.

Zhao M. and Barati M. [13] proposed an approach to detecting PPE in electrical substations based on graph neural networks (GNNs). Unlike classical object detection methods, which consider objects independently, the proposed model takes into account spatial and semantic relationships between objects, including employees, process equipment, and PPE, representing them as a graph structure. This approach improves the robustness and speed of PPE detection in a specific industrial environment characterized by a high density of objects, a complex background, and the presence of visually similar elements.

In the meantime, the authors note that the effectiveness of the proposed method largely depends on the quality and representativeness of the training data. Furthermore, constructing a graph structure requires additional labeling of the connections between scene objects, which increases the labour intensity of data preparation and complicates the implementation of such solutions in real-world industrial settings.

Ngoc-Thoan N., et al. [14] developed an improved model based on the YOLOv5 architecture for preventing emergency situations and monitoring the use of PPE at construction yards. The authors modified the network architecture and training procedure to improve the accuracy of detecting PPE elements and potentially hazardous industrial situations. Conducted experiments demonstrate improved performance compared to the base version of YOLOv5 model. Despite these results, the authors note that the developed model remains sensitive to changes in environmental conditions, including variations in illumination, weather conditions, and scene background characteristics. These factors may limit its reliability during around-the-clock operation in real industrial conditions.

Contemporary research is aimed at improving object detection architectures by eliminating anchoring mechanisms and improving the quality of data preparation. For example, Wang H. [15] investigated the applicability of the YOLOv8 model to PPE detection objectives. Based on the experimental results, the model demonstrates precision, recall, and F1-score values of 0.95, 0.84, and 0.74, respectively, outperforming models such as YOLOv5s, Faster R-CNN, SSD, and RetinaNet in detection quality.

YOLOv8's high performance is due to its use of an anchor-free detection mechanism and an expanded dataset using augmentation methods. Eliminating fixed anchors leads to the reduction of the model's dependence on predefined object parameters and to increasing its robustness to changes in scale, illumination, and scene dynamics. The results obtained demonstrate the high potential of the YOLOv8 architecture for building automated monitoring systems for PPE use in industrial conditions.

Materials and methods

Data collection and processing. A unique covering various scenarios and situations at production sites image dataset was formed for the study. For its formation, video materials obtained directly from the production sites were used. The initial sample included 300 video files, each of which was analyzed manually. Based on the results of the analysis, 91 of the highest quality video recordings were selected that met the requirements for training data. The selection took into account the absence of a black screen, sufficient lighting, the correct camera angle, the presence of activity in the frame and the minimum level of noise in the frame.

Based on the selected video materials, a dataset was created, including 2 877 images containing 13 object classes: person, helmet, glasses, jacket, gloves, pants, boots, as well as negative classes for each element of PPE (helmet-negative, glasses-negative, jacket-negative, gloves-negative, pants-negative, boots-negative). The use of negative classes made it possible to reduce the number of false positives in cases where elements of everyday clothing are visually similar to personal protective equipment.

Data Augmentation. To increase the volume and diversity of the dataset, augmentation methods were used, which is an important stage in data preparation in computer vision tasks. The use of data augmentation is relevant in the context of PPE detection, as it improves the model's ability to handle real-world scenarios.

During the augmentation process, geometric transformations were used, including rotation, scaling, and image translation, simulating changes in the position and orientation of objects typical of industrial settings. Horizontal image flipping was also applied, allowing the model to better generalize to PPE examples viewed from different angles. To simulate a variety of operating conditions, color corrections were performed on the images, including brightness and contrast adjustments. Blurring and noise addition methods were also applied to simulate real-world industrial conditions, where PPE may be partially hidden or exposed to external influences.

As a result of the augmentation, the dataset size increased from 2877 up to 16352 images. This allows for significant increase the variability of the training set and improvement of the model's robustness to changes in the external environment and shooting conditions.

Object Detection. Various computer vision architectures were considered for object detection and instance segmentation tasks, including Faster R-CNN [16], SSD [17], and models from the YOLO family [18]. Among them, special attention was given to the YOLOv8 architecture, developed by Ultralytics, which is an advanced single-stage detector optimized for real-time operation. The YOLOv8 architecture uses a fully convolutional approach with the CSPDarknet backbone, ensuring efficient gradient propagation and increasing computational efficiency. Unlike previous versions of the YOLO family, the model implements an anchor-free detection mechanism, which allows to increase the detector's flexibility and to reduce computational complexity without significantly reducing accuracy. For segmentation, YOLOv8 predicts masks simultaneously with bounding rectangles, using deformable convolutions to improve spatial perception of objects. By virtue of pre-training on large-scale datasets such as COCO, the model demonstrates high generalization ability and performance when solving a wide range of object detection tasks.

In this study, the YOLOv8 architecture was used to train and subsequently evaluate object detection and segmentation performance on the created dataset.

Human Pose Detection. Various deep learning architectures can be used to solve the problem of 2D human pose estimation, including ResNet [19], networks of Stacked Hourglass type [20], OpenPose [21], as well as methods based on Part Affinity Fields and Convolutional Pose Machines (CPMs) [22].

In this study, the model proposed by Sun K., et al. [23], based on the High-Resolution Network (HRNet) architecture was used to detect human body keypoints. HRNet uses a high-resolution subnet and repeatedly integrates representations formed at different resolutions, achieving high accuracy in pose estimation objective.

To predict keypoint coordinates in images, the HRNet model generates high-resolution heatmaps of objects, modeling the true position of each point using a two-dimensional Gaussian distribution. On the COCO and MPII datasets, the HRNet model demonstrates high accuracy in human pose estimation. This study used a pre-trained version of HRNet trained on the COCO dataset.

Model Training and Evaluation

Training the Object Segmentation Model. To train the object segmentation model, the resulting dataset was divided into subsets to effectively utilize its potential. The total dataset size was 16352 images, of which 13 081 were used for training the model, and the remaining 3 271 images were used for testing and validation. The distribution of objects by class is shown in Figure 1, allowing for a clear understanding of the dataset structure and its correspondence to real-world data.

The model was trained over 300 epochs, providing the model with a sufficient number of iterations to train on diverse data and adaptation to a variety of conditions.

Model Evaluation. To evaluate the performance of the YOLOv8 model for solving the PPE detection problem, standard object detection quality metrics were used: Precision, Recall, and Mean Average Precision (mAP). These metrics allow us to evaluate the model's ability to correctly detect and localize PPE in images.

Precision measures the accuracy of positive predictions made by the model. The formula for Precision is as follows:

$$Precision = \frac{TP}{TP+FP},$$

(1)

where *TP* (True Positives) is the number of correct PPE detections; *FP* (False Positives) is the number of instances where the model incorrectly predicted the presence of PPE despite its absence.

Recall, also known as sensitivity or true positive rate, measures the model's ability to detect all relevant PPE items. The formula for Recall is as follows:

$$Recall = \frac{TP}{TP+FN}$$

(2)

where *FN* (False Negatives) is the number of actual PPE items that the model failed to recognize.

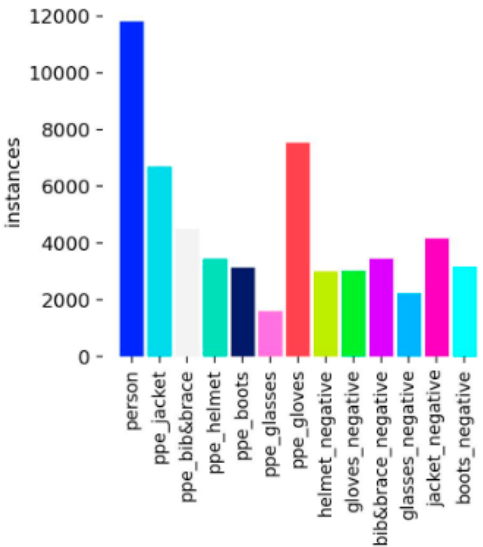


Figure 1. Class Distribution in the Dataset

Mean Average Precision (mAP) is one of the main metrics used to evaluate the quality of object detection models. This metric calculates the average accuracy at different confidence thresholds and is often visualized as a Precision-Recall Curve. Calculating mAP involves several steps:

- to calculate Precision and Recall for different confidence thresholds.
- to calculate the area under the Precision-Recall curve.
- to average the areas under the curve for different classes, which yields the mAP value.

The mAP provides a comprehensive picture of model performance at different confidence levels in detecting PPE objects. A higher mAP value indicates a more robust and accurate model.

The analysis of these metrics data allows to identify the strengths and weaknesses of the model, enabling adjustments to improve detection accuracy. To improve PPE detection accuracy, several key hyperparameters must be considered as follows: Learning rate is a critical parameter that affects model convergence. The optimal value is selected experimentally, starting with a low value and gradually increasing it if necessary.

Batch size affects stability and training speed. Larger batches speed up training but require more memory, so it is important to find a balance between efficiency and available resources.

Data augmentation involves the methods such as rotation, shear, brightness and contrast adjustments. improve the model's ability to generalize and adapt to real-world conditions.

Dataset balance ensures an even distribution of images from different classes (PPE and negative examples) to prevent the model from being biased toward a particular category.

A validation dataset was used to monitor training quality, playing a key role in assessing the model's generalization ability and identifying potential overfitting. The validation set was designed to cover various production scenarios, camera angles, and lighting conditions, ensuring its representativeness of real-world operating conditions. During validation, hyperparameters such as learning rate, batch size, and augmentation parameters were adjusted to determine the optimal configuration for the best object detection performance.

After completing the training and validation phases, the model was tested on a test dataset not previously used in the training process. The testing module evaluates the performance of the YOLOv8 model and verifies its accuracy in detecting PPE. For accurate evaluation, the test set must be diverse and cover potential operational scenarios in which the model will be used. The key testing metrics are Precision, Recall, and F1 score, which reflect the quality of predictions and the confidence of the model.

Therefore, when creating a YOLOv8 model, training parameters such as learning rate, batch size, and data augmentation methods must be considered. Balancing the dataset and ensuring diversity in the validation and testing datasets are crucial for accurate model evaluation. These recommendations should be adjusted to the specific requirements for improving PPE detection accuracy, ensuring the effective performance of the model.

For the further evaluation of the model's performance, the Precision-Confidence Curve was analyzed. This curve allows for evaluating the range of object detection accuracy changing with different model confidence thresholds. As shown in Figure 2, the model achieves an accuracy of approximately 0.98 for all classes, demonstrating high prediction reliability with a minimal number of false positives. It is of great importance in industrial conditions where strict compliance with safety standards is required.

The Recall-Confidence Curve (refer Figure 3) shows the relationship between Recall and model confidence. High Recall values, reaching 0.97, indicate the model's ability to effectively detect virtually all types of PPE, reducing the probability of missed events. This confirms the reliability of YOLOv8 for PPE detection tasks in real-world conditions.

An analysis of the Recall curves for individual classes reveals heterogeneity in the model's performance. Of particular interest is the "person" class, for which recall begins to decline sharply at a confidence threshold of approximately 0.4, while for most PPE classes, a steady decline is observed only at values above 0.7. This behavior is explained by the high

intra-class variability of objects in the "person" class, due to the variety of poses, viewing angles, object scales, and the presence of partial overlapping of employees in the frame. As a result, the model assigns lower confidence values to correct human detections.

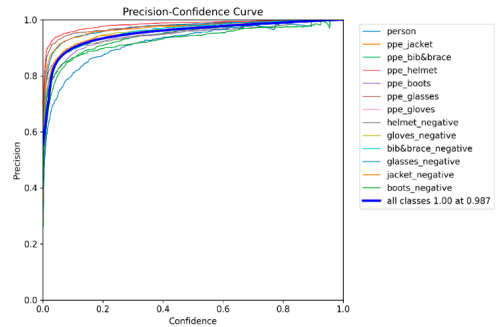


Figure 2. Precision-Confidence Curve based on bounding rectangles

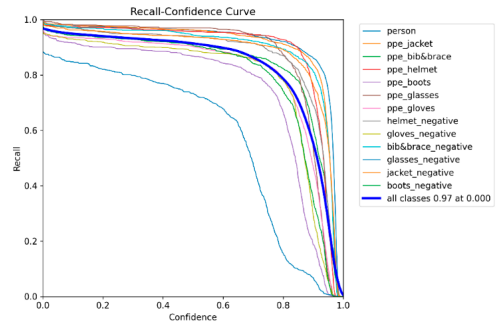


Figure 3. Recall-Confidence Curve based on bounding rectangles

For the comprehensive assessment of the quality of PPE detection, the Precision-Recall Curve was analyzed. Figure 4 illustrates the tradeoff between Precision and Recall, allowing for a detailed assessment of the model's performance. YOLOv8 demonstrates an average accuracy of $mAP@0.5$ equal to 0.964 across all object classes, demonstrating the model's high performance and reducing the risk of missing PPE items, contributing to improved safety in industrial conditions.

An analysis of the class-specific mean accuracy ($mAP@0.5$) values reveals significant variation: from 0.990 for the "person" class and 0.992 for "safety glasses" to 0.867 for "non-safety glasses" and 0.942 for "boots." The lowest value for the negative "non-safety glasses" class is conditional upon the visual proximity of regular eyeglasses and other facial features to safety glasses, making it difficult to distinguish between the positive and negative classes. The relatively low value for the "boots" class is conditional upon to the small size of the object

and the frequent overlap of shoes at the bottom of the frame when viewed from a wide-angle camera angle. These observations are consistent with the imbalanced distribution of classes in the dataset (Figure 1) and confirm the need for targeted expansion of the training set for poorly represented and visually ambiguous categories.

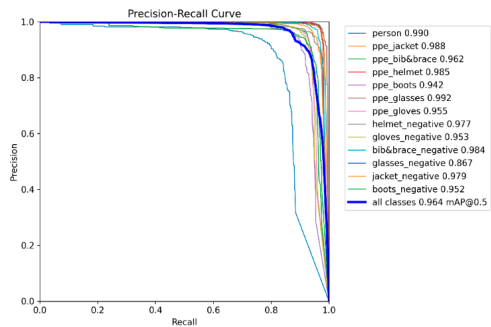


Figure 4. Precision-Recall curve for bounding rectangles

For the further evaluation of the model's quality, the F1-score dependence on confidence level curve (F1 Confidence Curve) was analyzed, allowing us to assess the balance between Precision and Recall of YOLOv8. Figure 5 demonstrates the influence of the confidence threshold on the F1-score value of the object detection quality metric. YOLOv8 achieves an F1-score of approximately 0.94 at the optimal confidence threshold for all PPE classes.

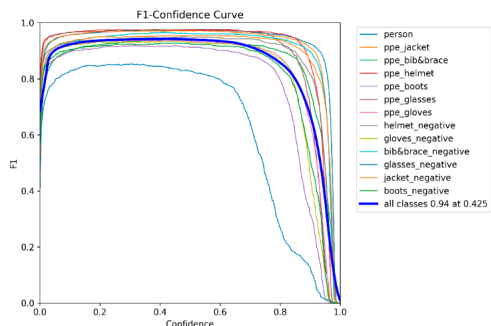


Figure 5. F1-score curve, confidence for bounding rectangles

The F1-score curve clearly shows that the optimal balance between Precision and Recall for all classes is achieved at a confidence threshold of approximately 0.425, resulting in an F1 value of 0.94. However, the “human” class demonstrates lower F1-score values (a plateau at approximately 0.8) across the entire threshold range compared to the PPE classes, which is consistent with the recall curve analysis. The model's results are presented in Fig. 6.

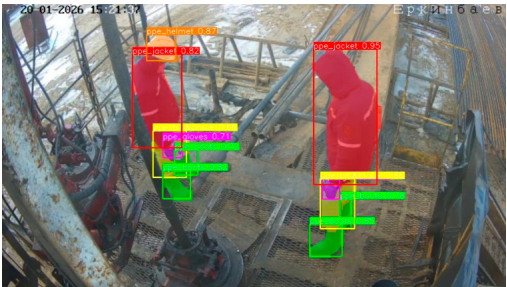


Figure 6. Model result on test data

Development of a PPE Monitoring System

To ensure continuous analysis of the video stream in real time, a dual-threaded data processing architecture is implemented in the system. The first thread (Capture Thread) asynchronously reads frames from the video stream and stores them in a buffer for subsequent processing. This approach minimizes delays in data receipt and ensures the transmission of the most relevant frames. The second thread (Processing Thread) is responsible for object detection and subsequent semantic analysis. It extracts frames from the first thread and processes them independently of the capture process. Splitting the functions into two threads reduces processing delays and ensures the relevance of the analyzed frames even when the computation time exceeds the interval between new images received. This approach prevents data transmission delays and efficiently utilizes computing resources (CPU/GPU).

At the first stage of video stream, processing a people detection model based on YOLOv8 and the BOTSORT tracker is used. These models effectively identify people in the frame and track their movements. Next, a pose estimation is performed for each detected person using the HRNet model. This model uses frame images and the coordinates of the person's bounding box (bbox) as input. The model can identify 17 key points of the human skeleton with high accuracy, which is especially important when multiple people are present in the frame.

Following the pose estimation, the person's bounding box is expanded by N pixels, after which the corresponding region of the image is cropped, and the cropped region is being uploaded to the object detection model. This approach allows for focused analysis on the employee location and eliminates the influence of the surrounding background. The model is trained to recognize 13 object classes: one class corresponds to the person, and the remaining 12 classes correspond to PPE items and their negative classes (helmets, goggles, gloves, etc.). Experiments have shown that using cropped images significantly improves the detection of small objects. For example, for the “gloves” class, 58 detections were obtained in cropped images,

compared to only 25 in the full frame. These results indicate that image cropping enhances the accuracy of small object detection, improving the overall performance of the system.

To reduce the number of false positives, a two-stage mechanism for matching detected objects to a specific employee was implemented. The first stage uses spatial matching, which matches the coordinates of detected PPE objects with the bounding rectangle of the person based on the intersection of their regions. The second stage uses anatomical matching, which verifies that the PPE matches the pose keypoints. For example, a helmet should be positioned near the head, and gloves near the hands.

The analysis results are stored in a "person" class object, which contains the following information: frames, a list of detected PPEs with confidence levels, a pose keypoint vector, and a tracking ID.

To improve decision stability, the system accumulates data for each person over M-consecutive frames ($M = 24$ at 8 FPS, which corresponds to a three-second video fragment). PPE presence classification is performed using the majority rule: if more than 50% of the frames in the interval contain a negative class (absence of an object), an event is generated. To reduce the impact of noise and random detection errors, a model confidence threshold exceeding 0.85 is additionally used for negative classes.

To prevent repeated registrations of the same violation, a time filter has been implemented. This filter uses a hash table of events with a set time-to-live (TTL) of 5 minutes. Each event is encoded with a combination of parameters (track ID, PPE type), preventing repeated registrations of the same violation within a specified time interval.

When a violation is detected in asynchronous mode, a GIF animation consisting of 24 consecutive frames is generated and then saved to the database via a separate record thread. This approach eliminates delays in the main thread associated with animation generation, maintaining high system performance.

Experiments have shown that generating GIF files is a resource-intensive operation and can reduce video stream processing speed. Therefore, the system implements a dynamic buffer clearing mechanism after object lifetimes, preventing the accumulation of out-dated data and optimizing RAM usage.

To improve system performance, TensorRT model quantization methods are used to accelerate GPU inference and an asynchronous frame-preloading algorithm. RAM consumption is controlled using an object manager, which removes excess instances of the Person class if their number exceeds a specified threshold (5 instances). This approach enables efficient resource management and ensures uninterrupted system operation even during lengthy video stream analysis.

Results and Discussion

This section presents the results of testing the PPE compliance monitoring system using 16 video

files obtained in real-world production conditions. The system uses two models: a PPE segmentation and detection model based on YOLOv8 and a human pose estimation model using HRNet, which significantly improves the accuracy of determining whether PPE is being worn correctly.

The following metrics were used to evaluate the model's performance: True Positives (TP) is equal to 113, True Negatives (TN) is equal to 0, False Positives (FP) is equal to 4, and False Negatives (FN) is equal to 52.

Based on these values, key metrics were calculated:

- Recall comes up to 0.6848. This metric characterizes the system's ability to detect actual violations, amounting to 68.48%. This indicates that the model is not always able to identify all violations;

- Precision comes up to 0.9658. A high value for this indicator indicates that 96.58% of the detected violations were correct;

- F1-score comes up to 0.8014. This metric is an integral indicator of the system's quality, as it takes into account both precision and recall.

- Accuracy comes up to 0.6686. This value indicates that 66.86% of the analyzed cases were correctly classified. However, this metric has limited informativeness for the task at hand, as the test sample did not contain any true negatives ($TN = 0$). This is due to the specifics of the dataset's formation: the "no violation" class was not included in the video tagging, as it has no practical significance in the context of the task. The model focuses exclusively on identifying violations, so only cases where they were present were considered during the tagging process. Consequently, no true negatives were recorded in the test videos.

The system demonstrated an F1 score of 0.8014, reflecting a balanced balance between high Precision (96.58%) and Recall (68.48%), as evidenced by a low false positive rate ($FP = 4$). This indicates a good ability of the model to detect violations. However, Recall (0.6848) indicates the presence of missed events ($FN = 52$). This suggests the need to further improve the system's sensitivity to reduce the number of undetected events.

To improve the completeness of violation detection in real-world conditions, it is necessary to optimize frame-processing algorithms, improve models for more accurate violation detection, and expand the training set.

Training vs. Real-World Gap. It is worth emphasizing the difference between the conditions under which the model is evaluated during training and during real-world use. Model training metrics (Precision = 0.98, Recall = 0.97) were obtained on a static test dataset containing 3,271 images with annotated objects, where detection quality was assessed exclusively under controlled conditions. The system training metrics (Precision = 0.9658, Recall = 0.6848) reflect the performance of the full monitoring system, which includes not only the detection model but also additional data processing stages: object tracking

(BOTSORT), temporal aggregation (24 frames in 3 seconds), threshold filtering (confidence > 0.85), anatomical matching, and a temporal filter for repeated events (TTL = 5 minutes). Each of these components can be a source of missed violations. Thus, the recall of 68.48% characterizes not an isolated detection model, but the full chain from video stream to event generation.

Comparative analysis with existing approaches. The obtained results should be considered in the context of previously published studies.

Table 1. Comparative analysis with existing approaches

Research	Model	Number of classes	Main results
Delhi V.S.K., Sankaral R., and Thomas A. [6]	YOLOv3	4	mAP = 97%, Recall = 97%, F1-score = 97%
Wang Z., et al. [7]	YOLOv5x	6	Accuracy = 86,55%
Lo J.-H., Lin L.-K., and Hung C.-C. [10]	YOLOv7	not specified	mAP = 97,29%
Wang H. [15]	YOLOv8	not specified	Precision = 0,95, Recall = 0,84, F1-score = 0,74
Model training (this study)	YOLOv8	13 (including 6 negative classes)	Precision = 0,98, Recall = 0,97, F1-score = 0,94
Digital monitoring of PPE (this study)	YOLOv8 + HRNet + BoT-SORT	13 (including 6 negative classes)	Precision = 0,9658, Recall = 0,6848, F1-score = 0,8014

As the table shows, during the training phase, the proposed YOLOv8 model demonstrates results comparable to the best solutions presented in the literature. However, when testing the monitoring system on real video recordings, Recall dropped to 68.48% while maintaining a high accuracy of 96.58%.

This discrepancy is explained by several significant factors. Firstly, the proposed system detects 13 object classes, including six negative classes, while most of the reviewed studies are limited to recognizing 2–6 object categories. Secondly, the system was evaluated on a continuous video stream obtained in real-world oil and gas field operational conditions, which presents additional challenges such as motion blur, changing illumination, partial occlusion of workers, and the small size of individual PPE elements in the frame.

Contribution of Anatomical Matching. The key difference between the proposed approach and works [5–7, 9–15] is the use of a two-stage matching mechanism, including spatial and anatomical matching. Most existing systems associate PPE with a worker based solely on the spatial intersection of bounding boxes (IoU), which can lead to false positives. For example, the system may mistakenly interpret a hard hat held by a worker as being correctly worn.

The closest approach to the proposed approach is the study by Xiong R. and Tang P. [4], which also uses human pose estimation to determine

attention zones. However, this approach is focused only on two types of PPE (hard hat and vest). The proposed system controls six types of PPE, including gloves and boots, which requires more precise determination of the positions of the hands and feet.

Using the HRNet architecture, it is possible to identify 17 key points of the human skeleton with high accuracy, ensuring correct anatomical matching of PPE elements with corresponding body parts even when multiple employees are present in the frame.

Practical Relevance and Industrial Implementation. Despite the moderate Recall value, the system's high accuracy (96.58%) ensures its practical significance for industrial applications. During the training phase, the YOLOv8 model achieved an F1 value of 0.94, significantly exceeding the result reported by Wang H. [15] for a similar YOLOv8s model (F1 = 0.74). Furthermore, the developed monitoring system, when processing a real video stream, demonstrated an F1 value of 0.8014, which is also higher than the results of the aforementioned study, despite significantly more complex operating conditions (13 classes, video stream, anatomical matching).

From a practical standpoint, high detection accuracy is of particular importance for industrial safety systems. Frequent false positives (unfounded warnings) lead to reduction of user confidence in the system and to ignoring the warnings by operators. Therefore, the achieved result, characterized by only four false positives (FP = 4) out of 117 positive predictions, demonstrates the system's high reliability and the accuracy of the generated violation notifications.

An additional advantage of the developed solution is its two-tier video stream processing architecture, based on asynchronous frame reading and processing. The implemented approach ensures a performance of 8 frames per second, which is sufficient for monitoring production processes characterized by relatively low employee movement speeds.

To improve computational efficiency, model quantization using TensorRT was used, which allows for reducing inference time and data processing delay. This solution is an important factor for the subsequent integration of the system into the existing video surveillance infrastructure of oil and gas fields and ensuring its operation in near-real time.

Limitations of the Research. Despite the obtained results, this study has several limitations. The system was tested on 16 video files obtained from a single oil field, which limits the generalizability of the results to other industrial facilities. The absence of a True Negatives class (TN = 0) is conditional upon to the nature of the dataset, which is designed solely for safety rules violation detection. Consequently, several quality indicators, such as Specificity and negative predictive value (NPV), cannot be calculated correctly. Furthermore, the dataset predominantly contains daylight scenes with a small

number of workers in the frame (up to 5 people), which may limit the system's effectiveness at nighttime shooting, complex lighting, and dense employee occurrence.

Directions for Further Research. To improve the comprehensiveness of safety rules violation detection, the following is planned:

1. To expand the dataset to more than 30 000 images, including scenes with night lighting, various weather conditions, and data from multiple industrial sites.
2. To explore adaptive mechanisms for selecting confidence thresholds instead of a fixed value (> 0.85), including the use of reinforcement learning methods for dynamic threshold adjustment.
3. To conduct a comparative evaluation of modern object detection models, including YOLOv11, YOLOv26, and RT-DETR, to improve the performance of small PPE elements detection.
4. To implement a feedback mechanism for further training the model based on accumulated false negatives (FNs).
5. To conduct multi-center validation of the system on data from various industrial sites to confirm the generalizability of the approach.

Conclusions

This paper presents the system for monitoring compliance with PPE wearing standards in real

time. The developed approach is based on a two-tier video stream processing architecture and combines the detection of 13 object classes using the YOLOv8 model, human pose estimation using HRNet, and a two-stage spatial and anatomical matching mechanism to verify proper PPE wearing.

Testing results on 16 video files obtained from a real oil field yielded an F1 score of 0.8014, with a Precision of 96.58% and a Recall of 68.48%. The system's high accuracy with a low false positives rate ($FP = 4$) confirms its practical applicability for industrial safety applications, where false positives reduce employee confidence in the system and ignoring warnings, creating greater operational damage than individual system gaps.

The key scientific and practical result of the study is the integration of anatomical matching based on human pose estimation, allowing the correlation of PPE elements with corresponding body parts. Unlike traditional approaches based solely on the spatial intersection of bounding rectangles, the proposed solution reduces the number of false positives and increases the reliability of safety compliance monitoring.

The results obtained confirm the potential of combining deep learning, object tracking, and human pose analysis for automated monitoring of PPE compliance.

ADDITIONAL INFORMATION

Funding source. This study was not supported by any external sources of funding.

Competing interests. The authors declare that they have no competing interests.

Authors' contribution. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work. The greatest contribution is distributed as follows: Galymzhan S. Abdimanap – data acquisition, initial data preprocessing, model training, and manuscript preparation; Anel N. Alimova – dataset annotation, manuscript drafting and article revision; Kairat A. Bostanbekov – model optimization; Renat Ye. Abdrakhmanov – image classification, data curation, model performance evaluation, and analysis of the obtained results; Dana Orenkyzy – image annotation; Daniyar B. Nurseitov – development of the research concept, dataset verification and validation, and interpretation of the obtained results.

ДОПОЛНИТЕЛЬНО

Источник финансирования. Авторы заявляют об отсутствии внешнего финансирования при проведении исследования.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Вклад авторов. Все авторы подтверждают соответствие своего авторства международным критериям ICMJE (все авторы внесли существенный вклад в разработку концепции, проведение исследования и подготовку статьи, прочли и одобрили финальную версию перед публикацией). Наибольший вклад распределён следующим образом: Эбдіманап F.C. – сбор и первичная обработка данных, обучение модели и написание статьи, Алимова A.H. – разметка датасета, подготовка текста и редакция статьи, Бостанбеков K.A. – оптимизация моделей, Абдрахманов P.E. – присвоение изображением классовых меток, систематизация данных, оценка эффективности модели и анализ полученных результатов, Оренқызы Д. – аннотирование изображений, Нурсейтов Д.Б. – разработка концепции исследования, проверка и валидация датасета, интерпретация полученных результатов.

REFERENCES

1. Kelm A, Laußat L, Meins-Becker A, et al. Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites. *Automation in Construction*. 2013;36:38–52. doi: [10.1016/j.autcon.2013.08.009](https://doi.org/10.1016/j.autcon.2013.08.009).

2. Zhang H, Yan X, Li H, et al. Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction. *Journal of Construction Engineering and Management*. 2019;145:1–13. doi: [10.1061/\(ASCE\)CO.1943-7862.0001629](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001629).
3. Dubey S, Dixit M. A comprehensive survey on human pose estimation approaches. *Multimedia Systems*. 2022;29:167–195. doi: [10.1007/s00530-022-00980-0](https://doi.org/10.1007/s00530-022-00980-0).
4. Xiong R, Tang P. Pose guided anchoring for detecting proper use of personal protective equipment. *Automation in Construction*. 2021;130:103828. doi: <https://doi.org/10.1016/j.autcon.2021.103828>.
5. Chen S, Demachi K. A vision-based approach for ensuring proper use of personal protective equipment (PPE) in decommissioning of Fukushima Daiichi nuclear power station. *Applied Sciences*. 2020;10(15):5129. doi: [10.3390/app10155129](https://doi.org/10.3390/app10155129).
6. Delhi VSK, Sankaral R, Thomas A. Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques. *Frontiers in Built Environment*. 2020;6:136. doi: [10.3389/fbuil.2020.00136](https://doi.org/10.3389/fbuil.2020.00136).
7. Wang Z, Wu Y, Yang L, et al. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches. *Sensors*. 2021;21(10):3478. doi: [10.3390/s21103478](https://doi.org/10.3390/s21103478).
8. GitHub [Internet]. ZijianWang Z. PPE_detection [cited 2026 May 05]. Available from: github.com/ZijianWang-ZW/PPE_detection.
9. Ma L, Li X, Dai X, et al. A Combined Detection Algorithm for Personal Protective Equipment Based on Lightweight YOLOv4 Model. *Wireless Communications and Mobile Computing*. 2022;3574588. doi: [10.1155/2022/3574588](https://doi.org/10.1155/2022/3574588).
10. Lo J-H, Lin L-K, Hung C-C. Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm. *Sustainability*. 2023;15(1):391. doi: [10.3390/su15010391](https://doi.org/10.3390/su15010391).
11. Lee Y-R, Jung S-H, Kang K-S, et al. Deep Learning-Based Framework for Monitoring Wearing Personal Protective Equipment on Construction Sites. *Journal of Computational Design and Engineering*. 2023;10(4):905–917. doi: [10.1093/jcde/qwad040](https://doi.org/10.1093/jcde/qwad040).
12. Nugraha KOPP, Rifai AP. Convolutional Neural Network for Identification of Personal Protective Equipment Usage Compliance in Manufacturing Laboratory. *Jurnal Ilmiah Teknik Industri*. 2023;22:11–24. doi: [10.23917/jiti.v22i1.21826](https://doi.org/10.23917/jiti.v22i1.21826).
13. Zhao M, Barati M. Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection Using GNN Approach. *IEEE Transactions on Industry Applications*. 2023;59(3):3142–3150. doi: [10.1109/TIA.2023.3234515](https://doi.org/10.1109/TIA.2023.3234515).
14. Ngoc-Thoan N, Bui D-QT, Tran CNN, Tran D-H. Improved Detection Network Model Based on YOLOv5 for Warning Safety in Construction Sites. *International Journal of Construction Management*. 2023;24(9):1007–1017. doi: [10.1080/15623599.2023.2171836](https://doi.org/10.1080/15623599.2023.2171836).
15. Wang H. Detection of Personal Protective Equipment (PPE) Using an Anchor Free-Convolutional Neural Network. *International Journal of Advanced Computer Science and Applications*. 2024;15(2):39. doi: [10.14569/IJACSA.2024.0150239](https://doi.org/10.14569/IJACSA.2024.0150239).
16. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;39(6):1137–1149. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
17. Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector. *Computer Vision – ECCV 2016*; 2016 Oct 11–14; Amsterdam, The Netherlands. 9905:21–37. Available from: link.springer.com/book/10.1007/978-3-319-46448-0.
18. Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conference on Computer Vision and Pattern Recognition*; 2016 June 27–30; Las Vegas, NV, USA. P:779–788. Available from: ieeexplore.ieee.org/document/7780460.
19. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*; 2016 June 27–30; Las Vegas, NV, USA. P:770–778. Available from: ieeexplore.ieee.org/document/7780459.
20. Newell A, Yang K, Deng J. Stacked Hourglass Networks for Human Pose Estimation Detector. *Computer Vision – ECCV 2016*; 2016 Oct 11–14; Amsterdam, The Netherlands. Vol. 9912. P:483–499. Available from: link.springer.com/book/10.1007/978-3-319-46484-8.
21. GitHub [Internet]. CMU-Perceptual-Computing-Lab/openpose [cited 2026 May 5]. Available from: <https://github.com/CMU-Perceptual-Computing-Lab/openpose>.
22. Wei S-E, Ramakrishna V, Kanade T, Sheikh Y. Convolutional Pose Machines. *IEEE Conference on Computer Vision and Pattern Recognition*; 2016 June 27–30; Las Vegas, NV, USA. P:4724–4732. Available from: ieeexplore.ieee.org/document/7780880.
23. Sun K, Xiao B, Liu D, Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 June 15–20; Long Beach, CA, USA. P:5686–5696. Available from: ieeexplore.ieee.org/document/8953615.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Kelm A., Laußat L., Meins-Becker A., et al. Mobile passive Radio Frequency Identification (RFID) portal for automated and rapid control of Personal Protective Equipment (PPE) on construction sites // *Automation in Construction*. 2013. Vol. 36. P. 38–52. doi: [10.1016/j.autcon.2013.08.009](https://doi.org/10.1016/j.autcon.2013.08.009).
2. Zhang H., Yan X., Li H., et al. Real-Time Alarming, Monitoring, and Locating for Non-Hard-Hat Use in Construction // *Journal of Construction Engineering and Management*. 2019. Vol. 145. P.1–13. doi: [10.1061/\(ASCE\)CO.1943-7862.0001629](https://doi.org/10.1061/(ASCE)CO.1943-7862.0001629).
3. Dubey S., Dixit M. A comprehensive survey on human pose estimation approaches // *Multimedia Systems*. 2022. Vol. 29. P. 167–195. doi: [10.1007/s00530-022-00980-0](https://doi.org/10.1007/s00530-022-00980-0).
4. Xiong R., Tang P. Pose guided anchoring for detecting proper use of personal protective equipment // *Automation in Construction*. 2021. Vol. 130. doi: [10.1016/j.autcon.2021.103828](https://doi.org/10.1016/j.autcon.2021.103828).
5. Chen S., Demachi K. A vision-based approach for ensuring proper use of personal protective equipment (PPE) in decommissioning of Fukushima Daiichi nuclear power station // *Applied Sciences*. 2020. Vol. 10, Issue 15. doi: [10.3390/app10155129](https://doi.org/10.3390/app10155129).
6. Delhi V.S.K., Sankaral R., Thomas A. Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques // *Frontiers in Built Environment*. 2020. Vol. 6. doi: [10.3389/fbuil.2020.00136](https://doi.org/10.3389/fbuil.2020.00136).
7. Wang Z., Wu Y., Yang L., et al. Fast Personal Protective Equipment Detection for Real Construction Sites Using Deep Learning Approaches // *Sensors*. 2021. Vol. 21, Issue 10. doi: [10.3390/s21103478](https://doi.org/10.3390/s21103478).

8. GitHub [интернет]. ZijianWang Z. PPE_detection [дата обращения 05.05.2026]. Доступ по ссылке: github.com/ZijianWang-ZW/PPE_detection.
9. Ma L., Li X., Dai X., et al. A Combined Detection Algorithm for Personal Protective Equipment Based on Lightweight YOLOv4 Model // *Wireless Communications and Mobile Computing*. 2022. doi: [10.1155/2022/3574588](https://doi.org/10.1155/2022/3574588).
10. Lo J.-H., Lin L.-K., Hung C.-C. Real-Time Personal Protective Equipment Compliance Detection Based on Deep Learning Algorithm // *Sustainability*. 2023. Vol. 15, Issue 1. doi: [10.3390/su15010391](https://doi.org/10.3390/su15010391).
11. Lee Y.-R., Jung S.-H., Kang K.-S., et al. Deep Learning-Based Framework for Monitoring Wearing Personal Protective Equipment on Construction Sites // *Journal of Computational Design and Engineering*. 2023. Vol. 10, Issue 4. P. 905–917. doi: [10.1093/jcde/qwad040](https://doi.org/10.1093/jcde/qwad040).
12. Nugraha Kh.O.P.P., Rifai A.P. Convolutional Neural Network for Identification of Personal Protective Equipment Usage Compliance in Manufacturing Laboratory // *Jurnal Ilmiah Teknik Industri*. 2023. Vol. 22. P. 11–24. doi: [10.23917/jiti.v22i1.21826](https://doi.org/10.23917/jiti.v22i1.21826).
13. Zhao M., Barati M. Substation Safety Awareness Intelligent Model: Fast Personal Protective Equipment Detection Using GNN Approach // *IEEE Transactions on Industry Applications*. 2023. Vol. 59, N 3. P. 3142–3150. doi: [10.1109/TIA.2023.3234515](https://doi.org/10.1109/TIA.2023.3234515).
14. Ngoc-Thoan N., Bui D.-Q. T., Tran C.N.N., Tran D.-H. Improved Detection Network Model Based on YOLOv5 for Warning Safety in Construction Sites // *International Journal of Construction Management*. 2023. Vol. 24, Issue 9. P. 1007–1017. doi: [10.1080/15623599.2023.2171836](https://doi.org/10.1080/15623599.2023.2171836).
15. Wang H. Detection of Personal Protective Equipment (PPE) Using an Anchor Free-Convolutional Neural Network // *International Journal of Advanced Computer Science and Applications*. 2024. Vol. 15, Issue 2. doi: [10.14569/IJACSA.2024.0150239](https://doi.org/10.14569/IJACSA.2024.0150239).
16. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. Vol. 39, N 6. P. 1137–1149. doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
17. Liu W., Anguelov D., Erhan D., et al. SSD: Single Shot MultiBox Detector // *Computer Vision – ECCV 2016*; October 11–14, 2016; Amsterdam, The Netherlands. Vol. 9905. P. 21–37. Available from: link.springer.com/book/10.1007/978-3-319-46448-0.
18. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection // *IEEE Conference on Computer Vision and Pattern Recognition*; June 27–30, 2016; Las Vegas, NV, USA. P. 779–788. Available from: ieeexplore.ieee.org/document/7780460.
19. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // *IEEE Conference on Computer Vision and Pattern Recognition*; June 27–30, 2016; Las Vegas, NV, USA. P. 770–778. Available from: ieeexplore.ieee.org/document/7780459.
20. Newell A., Yang K., Deng J. Stacked Hourglass Networks for Human Pose Estimation Detector // *Computer Vision – ECCV 2016*; October 11–14, 2016; Amsterdam, The Netherlands. Vol. 9912. P. 483–499. Available from: link.springer.com/book/10.1007/978-3-319-46484-8.
21. GitHub [интернет]. CMU-Perceptual-Computing-Lab/openpose [дата обращения 05.05.2026]. Доступ по ссылке: github.com/CMU-Perceptual-Computing-Lab/openpose.
22. Wei S.-E., Ramakrishna V., Kanade T., Sheikh Y. Convolutional Pose Machines // *IEEE Conference on Computer Vision and Pattern Recognition*; June 27–30, 2016; Las Vegas, NV, USA. P. 4724–4732. Available from: ieeexplore.ieee.org/document/7780880.
23. Sun K., Xiao B., Liu D., Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation // *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; June 15–20, 2019; Long Beach, CA, USA. P. 5686–5696. Available from: ieeexplore.ieee.org/document/8953615.

AUTHORS' INFO

Galymzhan S. Abdimanap

ORCID 0000-0003-1676-4075

e-mail: g.abdimanap@kmge.kz.

***Anel N. Alimova**

PhD

ORCID 0000-0002-5155-2417

e-mail: a.alimova@kmge.kz.

Kairat A. Bostanbekov

PhD

ORCID 0000-0003-2869-772X

e-mail: k.bostanbekov@kmge.kz.

Renat Ye. Abdrakhmanov

ORCID 0009-0004-1997-9818

e-mail: r.abdrakhmanov@kmge.kz.

Orenkyzy Dana

ORCID 0009-0000-2265-2091

e-mail: d.orenkyzy@kmge.kz.

Daniyar B. Nurseitov

канд. физ.-мат. наук, ассоциированный профессор

ORCID 0000-0003-1073-4254

e-mail: d.nurseitov@kmge.kz.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Әбдіманап Ғалымжан Сейтахметғалиұлы

ORCID 0000-0003-1676-4075

e-mail: g.abdimanap@kmge.kz.

***Алимова Анель Нурданбековна**

PhD

ORCID 0000-0002-5155-2417

e-mail: a.alimova@kmge.kz.

Бостанбеков Кайрат Аратович

PhD

ORCID 0000-0003-2869-772X

e-mail: k.bostanbekov@kmge.kz.

Абдрахманов Ренат Ермеқұлы

ORCID 0009-0004-1997-9818

e-mail: r.abdrakhmanov@kmge.kz.

Өренқызы Дана

ORCID 0009-0000-2265-2091

e-mail: d.orenkyzy@kmge.kz.

Нурсейитов Данияр Борисович

канд. физ.-мат. наук, ассоциированный профессор

ORCID 0000-0003-1073-4254

e-mail: d.nurseitov@kmge.kz.

*Автор, ответственный за переписку / Corresponding Author