

Ризванов Дмитрий Анварович

профессор кафедры вычислительной математики и кибернетики

Теплов Дмитрий Иванович

бакалавр, Уфимский университет науки и технологий, г. Уфа

РАСПОЗНАВАНИЕ ПОВЕДЕНИЯ СОТРУДНИКОВ НА РАБОЧИХ МЕСТАХ

Аннотация. В исследовании рассматривается задача автоматического распознавания действий сотрудников в офисе по видеопотоку с использованием свёрточной нейронной сети. Приведён анализ существующих методов детекции и классификации действий человека, описана разработанная архитектура нейросети на базе YOLOv8, представлены результаты экспериментального тестирования модели – полученная точность (mAP50) составила 86,1 процента. Работа предназначена для студентов, изучающих методы компьютерного зрения и машинного обучения, а также для специалистов в областях информационной безопасности, систем видеонаблюдения и искусственных нейронных сетей.

Annotation. This study examines the problem of automatic recognition of employee behavior in an office environment using video surveillance footage and convolutional neural networks. The paper provides an analysis of existing methods for human action detection and classification, describes the developed neural network architecture based on YOLOv8, and presents the results of experimental testing of the model – the achieved accuracy (mAP50) was 86.1 percent. The work is intended for students studying computer vision and machine learning methods, as well as for specialists in the fields of information security, video surveillance systems, and artificial neural networks.

Ключевые слова: методы машинного обучения, искусственный интеллект, офис, рабочие места, сотрудники офиса, компьютерное зрение, видеонаблюдение, свёрточные нейронные сети, многослойный персептрон, детекция объектов, распознавание поведения, классификация действий.

Keywords: machine learning methods, artificial intelligence, office, workplace, office employees, computer vision, video surveillance, convolutional neural networks, multilayer perceptron, object detection, behavior recognition, action classification.

Введение: современные системы видеонаблюдения позволяют фиксировать поведение сотрудников в офисах в течение всего рабочего дня. Однако ручной анализ многоканальных видеопотоков требует значительных временных и кадровых ресурсов, а также не позволяет оперативно выявлять нештатные ситуации (падения, длительное отсутствие на рабочем месте, конфликты). С развитием методов компьютерного зрения возникает необходимость внедрения автоматических систем распознавания действий сотрудников, способных анализировать видеопоток в реальном времени без участия человека и предоставлять количественную оценку поведения персонала.

Формулировка проблемы исследования: Термин «распознавание поведения сотрудников» (employee activity recognition) объединяет методы компьютерного зрения и машинного обучения, направленные на анализ действий человека по видеопотоку. Распознавание поведения – это метод автоматической классификации действий сотрудников, основанный на обработке видеоданных с камер наблюдения. Нейросетевая модель анализирует последовательность кадров, выделяя ключевые точки тела, позы и перемещения, и на основе этих данных классифицирует действие: сидит, стоит/ходит, разговаривает, падение.

Данная технология может быть использована для автоматизации контроля рабочего процесса и повышения безопасности в офисе. Так, системы распознавания поведения позволяют своевременно обнаруживать нештатные ситуации (падения, длительное отсутствие на рабочем месте), оценивать вовлечённость сотрудников в рабочий процесс и снижать нагрузку на службу безопасности. Несанкционированное использование систем видеонаблюдения без должного контроля может считаться нарушением приватности сотрудников [1Ошибка! Источник ссылки не найден.].

Известны случаи, когда отсутствие автоматического анализа видеопотока приводило к несвоевременному реагированию на падения сотрудников или конфликтные ситуации. Пострадавшие могли не получить своевременной помощи, потому что администратор не заметил инцидент на мониторах, а автоматическая система распознавания действий могла бы мгновенно оповестить службу безопасности [2][3].

Формулировка актуальности исследования: проблема автоматического контроля поведения сотрудников в офисе нуждается в комплексном решении. Сегодня, когда в большинстве офисных помещений установлены системы видеонаблюдения, возникают сложности с оперативным анализом многоканальных видеопотоков, так как ручной просмотр записей требует больших временных и кадровых затрат и не позволяет реагировать на события в реальном времени.

Руководители и службы безопасности особенно подвержены рискам, связанным с человеческим фактором, так как не всегда имеют возможность одновременно отслеживать все камеры наблюдения и своевременно выявлять нештатные ситуации. Сотрудник, оказавшийся в опасной ситуации (падение, потеря сознания, конфликт), может не получить своевременной помощи, если администратор не заметил инцидент на мониторах, а автоматическая система распознавания действий могла бы мгновенно оповестить службу безопасности.

В России по данным аналитических агентств в 2024 году рынок систем видеоаналитики вырос на 15–20%, при этом доля решений для распознавания поведения и нештатных ситуаций увеличилась с 5% до 12% от общего числа внедрений [4]. Это подтверждает рост востребованности автоматических систем распознавания действий сотрудников, способных работать в реальном времени и снижать нагрузку на человека.

Анализ существующих подходов: далее рассмотрим существующие подходы распознавания сгенерированных изображений:

1) Традиционные методы: обнаружение людей в кадре и отслеживание их перемещения на основе классических алгоритмов (вычитание фона, оптический поток, детекторы на основе HOG).

Методы анализируют движение объектов в кадре, например, изменение положения человека между кадрами, пересечение заданных зон, отклонение от типичных траекторий. Эти методы не используют машинное обучение, а основаны на ручных правилах и пороговых значениях.

2) Свёрточные нейронные сети: автоматическое извлечение пространственных признаков из изображений для выявления микро-артефактов, незаметных человеческому глазу.

Сверточные нейронные сети (например, предобученные «XceptionNet», «ResNeXt50», «MobileNet», «EfficientNet», «YOLO») обучаются на больших наборах данных, содержащих реальные и синтезированные изображения. Модель учится различать тонкие паттерны, как артефакты смешивания, нереалистичные текстуры кожи, искажения границ лица и другие.

3) Гибридные методы с анализом частотной области: комбинирование пространственных признаков (CNN) с временными характеристиками последовательности кадров.

Из последовательности кадров извлекаются пространственные признаки (позы, положение человека). Эти признаки подаются на вход рекуррентным сетям (LSTM, GRU) или трансформерам, которые анализируют динамику изменения позы во времени и классифицируют длительные действия.

4) Ансамблевые методы: комбинирование нескольких моделей или классификаторов для повышения точности и робастности детекции за счёт усреднения их результатов.

Несколько различных моделей (например, детектор объектов и классификатор поз) обучаются независимо. Затем их предсказания объединяются с помощью одного из методов: голосование, усреднение вероятностей или стекинг, когда предсказания базовых моделей подаются на вход мета-классификатору.

5) «Vision Transformers» для видео: использование архитектуры Transformer для захвата глобальных пространственно-временных зависимостей в видеопоследовательности.

Кадры разбиваются на патчи и подаются в Transformer, который моделирует взаимосвязи между патчами в пространстве и времени. Это позволяет одновременно учитывать локальные движения и глобальную структуру сцены. Механизм перекрёстного внимания объединяет пространственные и временные представления.

Ниже приведено сравнение рассмотренных подходов (таблица 1):

Таблица 1 – Сравнение подходов

Номер	Название	Точность (%)	Ссылка
1	Традиционные методы	60 – 75	[5]
2	Сверточные нейронные сети	60 – 95	[6]
3	Гибридные методы с анализом частотной области	90 – 95	[7]
4	Ансамблевые методы	95	[8]
5	«Vision Transformers»	98 – 99	[9]

Цель и задачи исследования: целью работы является разработка системы автоматического распознавания действий сотрудников в офисе по видеопотоку, позволяющая снять с человека обязанность по визуальному мониторингу видеокамер и обеспечить оперативное обнаружение нештатных ситуаций. В качестве основного инструмента такой системы предлагается использование свёрточной нейронной сети (CNN) на базе YOLOv8, обученной на размеченном наборе данных [10], содержащем кадры с классифицированными действиями: сидит, стоит/ходит, разговаривает, падение.

Для достижения поставленной цели необходимо решить следующие задачи:

- 1) Провести анализ предметной области:
 - 1.1) сформулировать проблему исследования;
 - 1.2) сформулировать актуальность исследования;

- 1.3) выполнить анализ существующих подходов.
- 2) Построить нейросетевую модель:
 - 2.1) подобрать данные для обучения модели;
 - 2.2) выбрать библиотеки для построения модели;
 - 2.3) подготовить данные для обучения;
 - 2.4) построить архитектуру модели;
 - 2.5) обучить модель;
 - 2.6) протестировать модель.
- 3) Осуществить анализ результатов:
 - 3.1) оценить качество модели;
 - 3.2) определить последствия внедрения предложенного решения для задач видеонаблюдения, безопасности и анализа продуктивности сотрудников.

Содержательное изложение подхода:

- 1) Структурная схема решения (рисунок 1):

№ слоя	<u>from</u>	<u>params</u>	Модуль	Аргументы
0	-1	928	Conv	[3, 32, 3, 2]
1	-1	18560	Conv	[32, 64, 3, 2]
2	-1	29056	C2f	[64, 64, 1, True]
3	-1	73984	Conv	[64, 128, 3, 2]
4	-1	197632	C2f	[128, 128, 2, True]
5	-1	295424	Conv	[128, 256, 3, 2]
6	-1	788480	C2f	[256, 256, 2, True]
7	-1	1180672	Conv	[256, 512, 3, 2]
8	-1	1838080	C2f	[512, 512, 1, True]
9	-1	656869	SPPF	[512, 512, 5]
10	-1	0	<u>Upsample</u>	[None, 2, 'nearest']
11	[-1, 6]	0	<u>Concat</u>	[1]
12	-1	591360	C2f	[768, 256, 1]
13	-1	0	<u>Upsample</u>	[None, 2, 'nearest']
14	[-1, 4]	0	<u>Concat</u>	[1]
15	-1	148224	C2f	[384, 128, 1]
16	-1	147712	Conv	[128, 128, 3, 2]
17	[-1, 12]	0	<u>Concat</u>	[1]
18	-1	493056	C2f	[384, 256, 1]
19	-1	590336	Conv	[256, 256, 3, 2]
20	[-1, -9]	0	Concat	[1]
21	-1	1969152	C2f	[768, 512, 1]
22	[15, 18, 21]	2117596	Detect	[4, 16, None, [128, 256, 512]]

Рисунок 1 – Структурная схема решения

- 2) Функциональная схема решения:



Рисунок 2 – Функциональная схема решения

3) Описание модели:

3.1) входной слой – для подачи изображения в модель;

3.2) скрытые слои (Архитектурные блоки Backbone): 10 последовательных блоков (Conv, C2F, SPPF) для извлечения пространственных признаков и наращивания каналов (функция активации *SiLU*);

3.3) скрытые слои (Блоки агрегации Neck): 12 блоков двунаправленного объединения признаков (Upsample, Concat) для комбинации с разных масштабов сети (функция активации *SiLU*);

3.4) 2 полносвязных слоя – для обработки признаков одномерного вектора;

3.5) выходной слой с операцией «sigmoid» – для формирования ответа.

4) Описание алгоритмов:

использованием «ImageDataLoader» осуществляется загрузка изображений из директорий (Train, Validation, Test) с автоматической нормализацией и пакетной обработкой, что позволяет оптимально расходовать оперативную память;

4.2) обучение – реализуется методом обратного распространения ошибки. На каждой эпохе модель последовательно обрабатывает пакеты изображений, вычисляет значение функции потерь, затем с помощью оптимизатора корректирует веса модели для минимизации ошибки. Процесс повторяется в течение 50 эпох;

4.3) тестирование – после завершения обучения модель фиксирует веса и выполняет классификацию изображений из тестовой выборки без дальнейшего изменения параметров. Результаты тестирования позволяют оценить обобщающую способность модели на ранее не встречавшихся данных.

5) Описание методов:

5.1) «AdamW» – адаптивный оптимизатор, который вычисляет индивидуальную скорость обучения для каждого параметра модели на основе оценок первого и второго моментов градиента;

5.2) «binary_crossentropy» – бинарная кросс-энтропия – функция потерь, используемая для задач бинарной классификации; измеряет различие между предсказанным распределением вероятностей и истинной меткой класса;

5.3) «mAP» – основная метрика для оценки качества детекции объектов; показывает среднюю точность модели при различных порогах уверенности и пересечении bounding boxes с реальными (IoU); используется для оценки качества обнаружения людей и их действий;

5.4) «precision» – метрика точности, показывающая, какая доля объектов, определённых моделью как «человек с действием X», действительно является таковыми, важна для снижения числа ложных срабатываний»;

5.5) «recall» – метрика полноты, показывающая, какая доля реальных объектов (людей с действием X) была правильно обнаружена моделью; важна для минимизации пропусков нештатных ситуаций (падений, длительного отсутствия).

6) Описание методик:

6.1) разбиение данных на тренировочную, валидационную и тестовую выборки;

6.2) нормализация данных — все изображения приводятся к единому масштабу путём деления значений пикселей (исходный диапазон $[0, 255]$) на 255, что преобразует их в диапазон $[0, 1]$. Такая нормализация ускоряет сходимость градиентного спуска и повышает устойчивость обучения, так как все входные признаки находятся в одном числовом диапазоне

Вычислительный эксперимент: полученные в ходе тестирования метрики ($mAP50 = 0,861$, $mAP50-95 = 0,619$, $precision = 0,884$, $recall = 0,892$) свидетельствуют о том, что модель успешно справляется с задачей распознавания действий сотрудников по видеопотоку. Высокое значение $precision$ означает низкий уровень ложных срабатываний — модель редко ошибочно классифицирует действие, когда сотрудник выполняет другое действие. Значение $recall$ показывает, что большинство реальных действий (особенно нештатных ситуаций, таких как падение) успешно обнаруживаются, что критически важно для систем безопасности.

Выводы: в ходе выполнения работы была достигнута поставленная цель — разработана система автоматического распознавания действий сотрудников в офисе по видеопотоку на основе свёрточной нейронной сети (YOLOv8). Решены все сформулированные задачи:

1) Проведён анализ предметной области: сформулирована проблема исследования, обоснована актуальность, выполнен анализ существующих подходов к распознаванию действий человека и готовых программных решений.

2) Построена нейросетевая модель: подобран датасет [**Ошибка! Источник ссылки не найден.Ошибка! Источник ссылки не найден.**], выбраны библиотеки (PyTorch, Ultralytics), разработана архитектура на базе YOLOv8, модель обучена и протестирована на валидационной выборке.

3) Осуществлён анализ результатов: получены метрики качества – модель правильно распознаёт действия сотрудников с точностью (mAP50) более 86%, при этом точность обнаружения нештатной ситуации (падение) составляет 89%, а полнота – 92%.

Практическая значимость разработанной системы заключается в возможности её интеграции в существующие системы видеонаблюдения офисных зданий для автоматического мониторинга действий сотрудников, своевременного обнаружения нештатных ситуаций (падение, длительное отсутствие на рабочем месте) и снижения нагрузки на службу безопасности, что позволяет повысить уровень безопасности и эффективности рабочего процесса.

Направления дальнейшего улучшения модели включают:

- увеличение количества эпох;
- добавление аугментации данных;
- использование предобученных архитектур;
- комбинирование пространственных признаков с частотным анализом.

Список литературы

1. Заржицкая Л. С. Использование работодателем систем видеонаблюдения: правовое регулирование и судебная практика // Мировой судья. 2024. № 9. С. 1–2. Электронная копия доступна в СПС «КонсультантПлюс». URL: <https://demo.consultant.ru/cgi/online.cgi?req=doc&base=CJI&n=154776> (дата обращения: 07.08.2026).

2. Видеоаналитика для охраны труда и промышленной безопасности. EYECONT // Малленом Системс: официальный сайт компании. URL: <https://www.mallenom.ru/resheniya/kontrol-lyudej/kontrol-prombezopasnosti/> (дата обращения: 07.08.2026).

3. И. М. Космачева, И. Ю. Кучин, Н. В. Давидюк [и др.] Система событийного мониторинга для автоматизированного обнаружения инцидентов // Вестник АГТУ. 2023. № 3. С. 76–86. URL: <https://cyberleninka.ru/article/n/sistema->

sobytiynogo-monitoringa-dlya-avtomatizirovannogo-obnaruzheniya-intsidentov (дата обращения: 07.08.2026).

4. Российский рынок видеонаблюдения и операционной аналитики [Электронный ресурс] // Ivideon. 2026. URL: <https://ru.ivideon.com/blog/bezopasnost-i-analitika-k-2030/> (дата обращения: 07.08.2026).

5. Захарова О. И., Бедняк С. Г., Шамрин Д. А. Сравнительный анализ методов машинного обучения с применением методов снижения размерности для классификации жестов рук // Прикаспийский журнал: управление и высокие технологии. 2024. № 1. С. 98–107. URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-metodov-mashinnogo-obucheniya-s-primeneniem-metodov-snizheniya-razmernosti-dlya-klassifikatsii-zhestov-ruk> (дата обращения: 07.08.2026).

6. Сайфутдинов А. В. Сверточные нейронные сети для решения задач компьютерного зрения // Universum: технические науки. 2023. № 10 (115). URL: <https://cyberleninka.ru/article/n/svertochnye-neyronnye-seti-dlya-resheniya-zadach-kompyuternogo-zreniya> (дата обращения: 07.08.2026).

7. Shahriar K. A., Ahmed M., Imtiaz H. Lightweight privacy-preserving human activity recognition from CSI data using a CNN-temporal attention network // Digital Discovery. 2026. Vol. 5, no. 6. P. 2701–2715. URL: <https://pubs.rsc.org/dd/article/5/6/2701/1226548/Lightweight-privacy-preserving-human-activity> (дата обращения: 07.08.2026).

8. Hsu T. W., Wu J., Liu H. Human activity recognition using an ensemble learning algorithm with smartphone sensor data // Electronics. 2022. Vol. 11, no. 3. Art. 322. URL: <https://www.mdpi.com/2079-9292/11/3/322> (дата обращения: 07.08.2026).

9. Li J., Yang Z. Human Action Recognition Using CNN and ViT // 2024 International Conference on Electronics and Devices, Computational Science (ICEDCS). IEEE, 2024. P. 107–111. URL: <https://www.computer.org/csdl/proceedings-article/icedcs/2024/276200a107/23yltjKS4qk> (дата обращения: 07.08.2026).

10. Edinburgh office monitoring video dataset [Electronic resource] / School of Informatics, University of Edinburgh. 2016. URL: <https://homepages.inf.ed.ac.uk/rbf/OFFICEDATA/> (дата обращения: 07.08.2026).

Resources

1. Zarzhitskaya L. S. Use of video surveillance systems by employers: legal regulation and judicial practice // Magistrate. 2024. No. 9. P. 1–2. Electronic copy available in the SPS "ConsultantPlus". URL: <https://demo.consultant.ru/cgi/online.cgi?req=doc&base=CJI&n=154776> (accessed: 07.08.2026).

2. Video analytics for occupational safety and industrial security. EYECONT // Mullenom Systems: official company website. URL: <https://www.mallenom.ru/resheniya/kontrol-lyudej/kontrol-prombezopasnosti/> (accessed: 07.08.2026).

3. Kosmacheva I. M., Kuchin I. Yu., Davidyuk N. V. [et al.] Event monitoring system for automated incident detection // Vestnik AGTU. 2023. No. 3. P. 76–86. URL: <https://cyberleninka.ru/article/n/sistema-sobytiynogo-monitoringa-dlya-avtomatizirovannogo-obnaruzheniya-intsidentov> (accessed: 07.08.2026).

4. Russian video surveillance and operational analytics market [Electronic resource] // Ivideon. 2026. URL: <https://ru.ivideon.com/blog/bezopasnost-i-analitika-k-2030/> (accessed: 07.08.2026).

5. Zakharova O. I., Bednyak S. G., Shamrin D. A. Comparative analysis of machine learning methods using dimensionality reduction techniques for hand gesture classification // Caspian Journal: Management and High Technologies. 2024. No. 1. P. 98–107. URL: <https://cyberleninka.ru/article/n/sravnitelnyy-analiz-metodov-mashinnogo-obucheniya-s-primeneniem-metodov-snizheniya-razmernosti-dlya-klassifikatsii-zhestov-ruk> (accessed: 07.08.2026).

6. Sayfutdinov A. V. Convolutional neural networks for solving computer vision problems // Universum: Technical Sciences. 2023. No. 10 (115). URL:

<https://cyberleninka.ru/article/n/svertochnye-neyronnye-seti-dlya-resheniya-zadach-kompyuternogo-zreniya> (accessed: 07.08.2026).

7. Shahriar K. A., Ahmed M., Imtiaz H. Lightweight privacy-preserving human activity recognition from CSI data using a CNN-temporal attention network // Digital Discovery. 2026. Vol. 5, no. 6. P. 2701–2715. URL: <https://pubs.rsc.org/dd/article/5/6/2701/1226548/Lightweight-privacy-preserving-human-activity> (accessed: 07.08.2026).

8. Hsu T. W., Wu J., Liu H. Human activity recognition using an ensemble learning algorithm with smartphone sensor data // Electronics. 2022. Vol. 11, no. 3. Art. 322. URL: <https://www.mdpi.com/2079-9292/11/3/322> (accessed: 07.08.2026).

9. Li J., Yang Z. Human Action Recognition Using CNN and ViT // 2024 International Conference on Electronics and Devices, Computational Science (ICEDCS). IEEE, 2024. P. 107–111. URL: <https://www.computer.org/csdl/proceedings-article/icedcs/2024/276200a107/23yltjKS4qk> (accessed: 07.08.2026).

10. Edinburgh office monitoring video dataset [Electronic resource] / School of Informatics, University of Edinburgh. 2016. URL: <https://homepages.inf.ed.ac.uk/rbf/OFFICEDATA/> (accessed: 07.08.2026).