

УДК 004.89

Глубокие нейронные сети в задаче распознавания агрессивного поведения на видео

Шведов Иван Александрович

ivan.shvedov7@mail.ru

Белов Юрий Сергеевич

ysbelov@bmstu.ru

КФ МГТУ им. Н.Э. Баумана, Калуга, Россия

Рассмотрено применение методов глубокого обучения для задачи распознавания агрессивного поведения на видеозаписях. Этот подход позволяет значительно повысить точность и надежность систем видеонаблюдения, автоматизируя процесс выявления потенциальных угроз. Глубокие нейронные сети, такие как сверточные модели (CNN), рекуррентные сети (RNN) и трансформеры, эффективно анализируют пространственно-временные зависимости в видеопотоке, что делает их незаменимыми в системах безопасности и мониторинга общественного порядка.

Ключевые слова: *глубокое обучение, распознавание, нейронные сети, поведение человека*

Глубокие нейронные сети (ГНС) представляют собой мощный инструмент для анализа видеоданных, включая задачу распознавания агрессивного поведения. Современные методы компьютерного зрения, основанные на глубоком обучении, позволяют извлекать сложные пространственно-временные зависимости из видеопотока, что особенно важно для детектирования действий в динамической среде. Распознавание агрессии на видео является актуальной задачей в таких областях, как системы безопасности, мониторинг общественного порядка, анализ поведения в медицине и психологии [1].

Одним из ключевых преимуществ ГНС в этой задаче является способность автоматически выделять релевантные признаки без необходимости ручного проектирования дескрипторов. Классические методы анализа видео, такие как оптический поток или гистограммы направлений градиентов, требуют явного задания признаков, что делает их менее универсальными по сравнению с глубокими моделями. Современные сверточные нейронные сети (CNN) и рекуррентные сети (RNN), включая их комбинации в виде архитектур CNN-LSTM, способны эффективно моделировать пространственные и временные зависимости, характерные для различных форм агрессивного поведения.

Основным источником данных для обучения таких систем являются размеченные видеодатасеты, содержащие примеры различных форм агрессии, например драки, крики, угрожающие жесты. Среди популярных датасетов можно выделить Hockey Fight, Movies Fight Dataset, UCF Crime и другие, содержащие как постановочные, так и реальные сцены агрессии. Для обучения глубокой нейронной сети обычно требуется большой объем данных, поэтому активно применяются методы увеличения данных (data augmentation), такие

как изменение яркости, вращение, отражение кадров, а также генерация синтетических примеров с помощью генеративных моделей [2].

Существуют различные подходы к построению моделей для детектирования агрессии. Один из наиболее распространенных — использование предварительно обученных сверточных сетей, таких как ResNet, EfficientNet или MobileNet, в качестве экстракторов признаков, которые затем передаются в рекуррентную сеть, например LSTM или GRU, для обработки временной информации. Такой подход позволяет комбинировать мощные возможности CNN по выделению пространственных особенностей с возможностью RNN анализировать последовательность кадров. Более сложные архитектуры, такие как трансформеры, включая Vision Transformer (ViT) и его модификации, также начинают активно применяться в этой задаче, поскольку способны эффективно моделировать долгосрочные зависимости в видеопотоке.

Одним из важных аспектов является выбор метода детектирования агрессивного поведения. Подходы делятся на кадровый и потоковый анализ. В первом случае модель оценивает каждый кадр видео независимо, что требует постобработки результатов для выявления последовательностей с агрессивными действиями. Потоковый анализ, напротив, учитывает всю последовательность кадров сразу, что делает его более точным, но требует значительных вычислительных ресурсов. Важную роль играют методы постобработки, включая фильтрацию ложных срабатываний, объединение фрагментов с агрессией в единый временной сегмент и расчет вероятности агрессивного поведения на основе нескольких источников данных.

Эффективность работы ГНС в задаче распознавания агрессивного поведения на видео во многом зависит от качества разметки данных и стратегии обучения. Один из популярных методов повышения точности — использование многозадачного обучения (multi-task learning), когда сеть обучается одновременно на нескольких связанных задачах, например детекции объектов, анализе поз и предсказании действия. Также применяются методы самонаблюдаемого обучения (self-supervised learning), позволяющие использовать неразмеченные данные для предварительного обучения модели перед ее дообучением на размеченном датасете [3].

На практике системы автоматического анализа агрессии на видео сталкиваются с рядом сложностей. К ним относятся вариативность условий съемки, разнообразие агрессивных действий, индивидуальные различия в поведении людей, а также возможные искажения видео, такие как шум, низкое разрешение или сильное освещение. Для повышения надежности используются ансамблевые методы, объединяющие предсказания нескольких моделей, а также байесовские подходы для оценки степени уверенности предсказания.

В последние годы активно развиваются методы объяснимого искусственного интеллекта (XAI), позволяющие понять, какие именно признаки использует нейронная сеть для классификации агрессивного поведения. Такие методы, как Grad-CAM и SHAP, помогают выявить ключевые области изображения, на которые ориентируется модель, что способствует ее дора-

ботке и адаптации к новым условиям. Это особенно важно в контексте практического применения, например в судебных разбирательствах или системах мониторинга общественного порядка, где требуется объяснимость решений модели [4].

Таким образом, глубокие нейронные сети демонстрируют высокую эффективность в задаче распознавания агрессивного поведения на видео, позволяя автоматизировать процесс анализа и снизить нагрузку на операторов систем безопасности. Однако для достижения высокой точности требуется тщательная подготовка данных, использование современных архитектур и учет реальных условий применения. Развитие новых методов, таких как трансформеры, самообучение и объяснимый искусственный интеллект, будет способствовать дальнейшему улучшению качества детекции агрессии, делая системы более надежными и адаптивными к различным сценариям.

Литература

- [1] Zhang H.B., Zhang Y.X., Zhong B., Lei Q., Yang L., Du J.X., Chen D.S. A comprehensive survey of vision-based human action recognition methods. *Sensors*, 2019, vol. 19 (5), no. 1005.
- [2] Sujata Gupta N., Ramya K., Karnati R. A Review Work: Human action recognition in video surveillance using deep learning techniques. *Informatics and Automation*, 2024, vol. 23, no. 2, pp. 436–466. <https://doi.org/10.15622/ia.23.2.5>
- [3] Zhang J., Zi L., Hou Y., Wang M., Jiang W., Deng D. A DL-based approach to enable action recognition for construction equipment. *Advances in Civil Engineering*, 2020, pp. 1–14. <https://doi.org/10.1155/2020/8812928>
- [4] Козловский А.В., Волощук В.И., Мельник Я.Э. Система распознавания девиантного поведения в видеопотоке с использованием нейронных сетей и других методов машинного обучения. *Известия ТулГУ. Технические науки*, 2023, № 2, с. 77–82.

Deep neural networks in the task of recognizing aggressive behavior on video

Shvedov Ivan Alexandrovich

ivan.shvedov7@mail.ru

Belov Yuri Sergeevich

ysbelov@bmstu.ru

Kaluga Branch of BMSTU, Kaluga, Russia

The article discusses the application of deep learning methods for the task of recognizing aggressive behavior on video recordings. This approach significantly improves the accuracy and reliability of video surveillance systems by automating the process of identifying potential threats. Deep neural networks, such as convolutional models (CNNs), recurrent networks (RNNs), and transformers, effectively analyze the space-time dependencies in a video stream, which makes them indispensable in security and public order monitoring systems.

Keywords: deep learning, recognition, neural networks, human behavior