

УДК 004.932.2

ИЕРАРХИЧЕСКАЯ ВИДЕОКЛАССИФИКАЦИЯ НЕБЕЗОПАСНЫХ ПРОИЗВОДСТВЕННЫХ СИТУАЦИЙ

А. А. АРХИПОВ, М. А. ДУДИН, И. И. МАКСИМОВ

Научный руководитель А. Д. ОБУХОВ, д-р техн. наук, проф.

Тамбовский государственный технический университет
Тамбов, Россия

Распознавание нарушений промышленной безопасности требует учитывать позу человека, состояние оборудования, разметку прохода и развитие действия во времени [1, 2]. Покадровая классификация теряет временной контекст, а малый несбалансированный датасет повышает вероятность переобучения. Целью эксперимента являлась разработка модели, одновременно определяющей ситуацию, её семейство и бинарный признак безопасности.

В качестве основы выбран VideoMAE Base, предварительно обученный на Kinetics [2]. Модель получает 16 равномерно выбранных RGB-кадров полного плана размером 224×224 . Сохранение полного кадра принципиально важно: классы открытой панели, нарушения маршрута и перегрузки погрузчика различаются по контексту, который может отсутствовать в стор человека. Усреднённое представление backbone поступает в три классификационные головы: восемь исходных классов, четыре семейства ситуаций и два класса safe/unsafe.

Таксономия образована четырьмя парами: безопасный маршрут и его нарушение; разрешённое и несанкционированное вмешательство; закрытая и открытая панель; безопасная транспортировка и перегрузка погрузчика. Функция потерь объединяет cross-entropy детальной головы с весом 1,0, семейства – 0,35, безопасности – 0,45 и KL-согласованность – 0,15. Дисбаланс компенсируют взвешенная выборка, веса классов и focal-множитель [3].

Датасет содержит 691 клип: 566 в обучающей и 125 в независимой тестовой части. Число примеров класса изменяется от 30 до 210. Из train стратифицированно выделено 15 % для validation. Официальный test не участвовал в подборе параметров. Обучение проводилось на CUDA 25 эпох с early stopping по macro F1. Batch size равен 4, gradient accumulation – 4, learning rate backbone – $3 \cdot 10^{-6}$, голов – $5 \cdot 10^{-5}$. Использовались BF16, AdamW, focal gamma 1,5 и label smoothing 0,03 (табл. 1).

Табл. 1. Результаты оценки иерархической VideoMAE

Метрика	Значение
Лучшая validation macro F1	0,652
Test macro F1	0,643
Test balanced accuracy	0,665
Unsafe recall	0,875
Unsafe precision	0,727

Наибольший F1 получен для безопасной транспортировки и перегрузки погрузчика – 1,0 при восьми тестовых примерах каждого класса. Для нарушения маршрута F1 составил 0,655, для несанкционированного вмешательства – 0,552. Наиболее сложны разрешенное вмешательство (0,455) и безопасный маршрут (0,400), что объясняется малым числом примеров и сходством парных ситуаций.

Значение unsafe recall 0,875 показывает, что модель обнаруживает основную часть небезопасных клипов. Для системы охраны труда эта характеристика приоритетнее общей точности. Unsafe precision 0,727 указывает на наличие ложных тревог, которые в разработанной системе компенсируются независимыми правилами, дополнительной VLM-проверкой evidence-кадра и решением оператора. В online-контур экспортирована лучшая одиночная модель. Прогноз выполняется по скользящему окну из 16 кадров с задаваемым шагом и порогами confidence 0,55 и unsafe probability 0,5.

Эксперимент подтвердил применимость иерархической VideoMAE к малому промышленному датасету. Дальнейшее повышение качества связано с пополнением редких классов, grouped-разбиением по эпизодам, hard-negative mining и калибровкой unsafe-порога. Для промышленного исполнения целесообразны экспорт в ONNX/TensorRT и оценка задержки на целевом оборудовании.

Работа выполнена при финансовой поддержке Министерства образования и науки Тамбовской области в рамках проекта № МУ2025-02/23 областного конкурса «Гранты для поддержки прикладных научных исследований молодых учёных в 2025 году».

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. **Николюкин, М. С.** Применение компьютерного зрения для мониторинга и анализа процессов на промышленных объектах / М. С. Николюкин, Н. С. Недосып, А. Е. Архипов // Виртуальное моделирование, прототипирование и промышленный дизайн : материалы X Междунар. науч.-практ. конф., Тамбов, 21–22 окт. 2024 г. – Тамбов: Тамбов. гос. техн. ун-т, 2024. – С. 249–251.
2. **Круглов, В. С.** Разработка мультиплатформенной системы классификации упражнений для опорно-двигательной реабилитации / В. С. Круглов, Ю. В. Никитников, А. Д. Обухов // Виртуальное моделирование, прототипирование и промышленный дизайн : материалы XI Междунар. науч.-практ. конф., Тамбов, 10–12 нояб. 2025 г. – Тамбов : Тамбов. гос. техн. ун-т, 2025. – С. 160–162.
3. VideoMAE: Masked Autoencoders are Data-Efficient Learners for Self-Supervised Video Pre-Training / Z. Tong, Y. Song, J. Wang, L. Wang // Advances in Neural Information Processing Systems. – 2022. – Vol. 35. – P. 10078–10093.
4. Focal Loss for Dense Object Detection / T.-Y. Lin, P. Goyal, R. Girshick [et al.] // Proceedings of the IEEE International Conference on Computer Vision. – 2017. – P. 2980–2988.