УДК 004.932 : 615.47

ОЦЕНКА КРОСС-ДОМЕННОГО ОБОБЩЕНИЯ СОВРЕМЕННЫХ НЕЙРОАРХИТЕКТУР В ЗАДАЧЕ ДЕТЕКЦИИ ПАДЕНИЙ ПО СКЕЛЕТНЫМ КЛЮЧЕВЫМ ТОЧКАМ

Д.А. Ермолаев[1], А.С. Кройдер[1], Ф.А. Белянин[2]

[1]Самарский университет, г. Самара
[2]Самара секьюрити сервис, г. Самара
e-mail: danilssaustudent5104@gmail.com

Аннотация

В работе рассматривается возможность использования нейросетевых моделей для классификации падения человека по извлеченным ключевым координатам позы с помощью кросс-доменного обобщения. В качестве классификаторов были использованы современные архитектуры: LSTM, GRU, Transformer Encoder и 1D-CNN + GAP. Модели обучались и тестировались на датасете CAUCAFall. На независимом датасете FallVision без дообучения оценено кросс-доменное обобщение. Чтобы подтвердить статистическую значимость полученных результатов на обоих датасетах была проведена проверка значимости с помощью тестов McNemar, bootstrap-доверительные интервалы для F1, перестановочный тест. Показано, что высокие метрики на исходном датасете статистически значимы. При тестировании моделей на независимом датасете без обучения наилучший баланс качества и сложности демонстрирует 1D-CNN + GAP при существенно меньшем числе параметров, что важно для внедрения на ограниченных по ресурсам устройствах.

Ключевые слова: нейросетевые модели, детекция падений, кросс-доменный перенос, статистическая значимость.

Введение

По данным ВОЗ, для лиц старше 65 лет ежегодная частота падений составляет 28–35 %, увеличиваясь до 32–42 % у лиц старше 70 лет [1]. Для мониторинга падений перспективны системы с локальной обработкой на edge-камерах. Широко используется двухэтапный метод: формирование скелетного представления по ключевым точкам, а затем бинарная классификация представлений. Такой метод уменьшает зависимость от внешних факторов и позволяет внедрять технолгию в ресурсоограниченные устройства [2].

Цель работы – оценить способность модели сохранять высокую точность при работе на новых данных и статистически проверить качество детекции.

Материалы и методы

В данной работе использовались CAUCAFall [4] и FallVision [5] датасеты. Оба содержат записи падений в домашних условиях. CAUCAFall включает 10 участников, имеет частоту кадров 23 fps и 5 сценариев падений из положения сидя и стоя. FallVision значительно разнообразнее по составу (58 участников), снимался с частотой 30 fps,

включает падения из положения стоя, сидя, а также с кровати, причём последняя категория отсутствует в обучающей выборке.

Для проведения эксперимента были выбраны следующие виды современных архитектурных решений для классификации:

1. Long ShortTerm Memory (LSTM) [6] и Gated Recurrent Unit (GRU) [5] являются рекуррентными моделями, состоящие из двух слоев и извлекающие долгосрочные и краткосрочные зависимости из последовательностей.

2. Архитектура Transformer Encoder [7] основана на механизме самовнимания, анализирующая глобальный контекст, что важно для детекции меняющихся во времени собитий.

3. One Dimensional-Convolution Neural Network с Global Average Pooling (1D-CNN + GAP) [8] представляет собой одномерные свёртки по оси времени с пулингом и глобальным усреднением.

Для оценки качества распознавания использовались Accuracy, F1-мера и AUC-ROC. Для подтверждения статистической надежности результатов были применены тест Мак-Немара [9] и пермутационный тест [10]. Критерий Мак-Немара оценивает, есть ли значимое различие между двумя моделями, сравнивая только те примеры, где их ответы не совпали. Пермутационный тест был применен для проверки значимости результатов каждой модели относительно случайного угадывания на тестовых данных. Суть метода заключается в многократном случайном перемешивании меток классов и пересчете метрики F1.

1. Результаты и обсуждения

В таблице 2 представлены статистически проверенные результаты тестирования моделей.

Таблица 2 – Статистические подтвержденные результаты моделей на датасетах CAUCAFall и FallVision

Модель	Параметры	F1	AUC-ROC	Acc	Пермутационный тест	Мак-Немара
CAUCAFall						
LSTM	225 282	0,984	0,9995	0,987	< 0,0001	-
GRU	169 090	0,961	0,9985	0,968	< 0,0001	0,248
Transformer	403 714	0,968	0,9988	0,975	< 0,0001	0,617
1D-CNN + GAP	69 762	0,976	0,9987	0,981	< 0,0001	1,000
FallVision						
LSTM	225 282	0,709	0,792	0,720	< 0,0001	-
GRU	169 090	0,670	0,704	0,778	< 0,0001	< 0,001
Transformer	403 714	0,655	0,793	0,699	< 0,0001	< 0,001
1D-CNN + GAP	69 762	0,714	0,799	0,725	< 0,0001	0,214

При кросс-доменном тестировании на независимом датасете FallVision наблюдается заметное снижение всех ключевых метрик по сравнению с результатами на CAUCAFall. F1-мера падает с 0,96–0,98 до 0,65–0,71, Accuracy – с 0,97–0,99 до

0,70–0,78, а AUC-ROC – с 0,999 до 0,70–0,80. Это подтверждает существенные доменные различия и подчёркивает важность оценки обобщения на реальных данных.

Анализ данных пермутационного теста для всех архитектур на двух датасетах статистически доказывает, что модели извлекают реальные кинематические паттерны из видеопотока, а не подстраиваются под распределение классов.

Несмотря на то, что сверточная модель 1D-CNN + GAP содержит в 3,2 раза меньше параметров, чем классическая рекуррентная сеть LSTM, тест Мак-Немара показал значение p = 1,0 и p = 0,214 на первом и втором датасетах соответственно, что подтверждает разницу в качестве предсказаний между этими моделями статистически не значима. В то же время более емкие архитектуры GRU и Transformer показали значимое снижение точности при кросс-доменном переносе. Это подтверждает гипотезу о том, что избыточная сложность моделей может приводить к переобучению, тогда как одномерные свертки лучше выделяют инвариантные признаки.

Заключение

Статистическая эквивалентность модели 1D-CNN + GAP в сравнении с более сложными аналогами делает её оптимальным выбором для интеграции в edge-устройства. Статистические результаты показывают, что высокие метрики обусловлены не случайным уровнем классификации.

Список использованных источников

1. World Health Organization. WHO global report on falls prevention in older age. – Geneva: World Health Organization, 2007. – URL: https://www.who.int/publications/i/item/9789241563536 (дата обращения: 20.03.2026).

2. Fall Detection and Activity Recognition Using Human Skeleton Joints / H. Ramirez [et al.] // IEEE Access. – 2021. – V. 9. – P. 9360751.

3. Dataset for human fall recognition in an uncontrolled environment / J. C. E. Guerrero [et al.] // Data in Brief. – 2022. – V. 44. – P. 108522.

4. FallVision: A benchmark video dataset for fall detection / N. N. Rahman [et al.] // Data in Brief. – 2025. – V. 68. – P. 108000.

5. Fall Detection Based on Human Skeleton Keypoints Using GRU / Y. K. Kang, H. Y. Kang [et al.] // International Journal of Internet. – 2020. – V. 12, № 4. – P. 83–92.

6. A novel vision-based fall detection scheme using keypoints of human skeleton with long short-term memory network / A. R. Inturi [et al.] // Arabian Journal for Science and Engineering. – 2023. – V. 48, № 2. – P. 1143–1155.

7. Computer Vision Based Transfer Learning-Aided Transformer Model for Fall Detection // IEEE JBHI, 2024. – T. 28, № 6.

8. Rethinking 1D-CNN for Time Series Classification: A Stronger Baseline / W. Tang [et al.] // arXiv preprint arXiv:2002.08561. – 2020. – P. 113–115

9. McNemar, Q. Note on the sampling error of the difference between correlated proportions or percentages / Q. McNemar // Psychometrika. – 1947. –№ 2. – P. 153–157.

10. Ojala, M. Permutation Tests for Studying Classifier Performance / M. Ojala, G. C. Garriga // Journal of Machine Learning Research. – 2010. – V. 11. – P. 1833–1863.

EVALUATION OF CROSS-DOMAIN GENERALIZATION OF MODERN NEURAL ARCHITECTURES IN THE TASK OF FALL DETECTION BASED ON SKELETAL KEYPOINTS

D.A. Ermolaev[1], A.S. Kroyder[1], F.A. Belyanin[2]

[1]Samara University, Samara;
[2]Samara security services, Samara
e-mail: danilssaustudent5104@gmail.com

Abstract

The paper examines the possibility of using neural network models for human fall classification based on extracted key pose coordinates via cross-domain generalization. Modern architectures such as LSTM, GRU, Transformer Encoder, and 1D-CNN + GAP were employed as classifiers. The models were trained and tested on the CAUCAFall dataset. Cross-domain generalization was evaluated on the independent FallVision dataset without additional training. Statistical significance of the results on both datasets was verified using McNemar test, bootstrap confidence intervals for F1-score, and permutation test.

It is demonstrated that the high metrics on the original dataset are statistically significant. When tested on the independent dataset without retraining, the 1D-CNN + GAP model provides the best balance of quality and complexity with a substantially smaller number of parameters, which is crucial for deployment on resource-constrained devices.

Keywords: neural network models, fall detection, cross-domain transfer, statistical significance.