Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ Информатика, искусственный интеллект и системы управления
КАФЕДРА Системы обработки информации и управления

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

Модели распознавания движений по данным видеонаблюдения

Студент ИУ5-34М _____ **Е.И. Мащенко**
(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель _____ **Ю.Е. Гапанюк.**
(Подпись, дата) (И.О.Фамилия)

2023 г.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ

Заведующий кафедрой _ИУ5_____
(Индекс)

_____ _____ _____Терехов В.И__
(И.О.Фамилия)

« _____ » _____ 20 ____ г.

З А Д А Н И Е
на выполнение научно-исследовательской работы

по теме ___Модели распознавания движений по данным видеонаблюдения_____
_____

Студент группы ___ИУ5-34М_____

_____Мащенко Елена Игоревна_____
(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

_____исследовательская_____

Источник тематики (кафедра, предприятие, НИР) _____кафедра_____

График выполнения НИР: 25% к _4_ нед., 50% к _8_ нед., 75% к _12 нед., 100% к _16_ нед.

Техническое задание ___Проведение исследования по теме _____
_____
_____
_____

Оформление научно-исследовательской работы:

Расчетно-пояснительная записка на _19___ листах формата А4.
Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

_____
_____
_____
_____

Дата выдачи задания « ___ » _____ 20__ г.

Руководитель НИР _____ _Гапанюк Ю.Е._____
(Подпись, дата) (И.О.Фамилия)

Студент _____ __Мащенко Е.И._____
(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Одной из самых серьезных задач в современном обществе является улучшение условий здорового старения и поддержка пожилых людей в их повседневной деятельности. В частности, автоматическое обнаружение падений привлекло значительное внимание в сообществах, занимающихся компьютерным зрением и распознаванием образов.

Хотя подходы, основанные на носимых датчиках, обеспечили высокие показатели обнаружения, некоторые потенциальные пользователи неохотно их носят, и, таким образом, их использование еще не нормализовано. Как следствие, появились альтернативные подходы, такие как методы, основанные на компьютерном зрении. Внедрение парадигм "умной среды" и Интернета вещей, а также растущее число камер в нашей повседневной среде создают оптимальный контекст для систем, основанных на зрении.

Целью работы является проведение исследования по теме обнаружения падений по данным видеонаблюдения.

В предыдущей работе была разработана модель сверточной нейронной сети, определяющая по изображению, упал человек или нет. В данной работе продолжаются исследования по данной теме и создается модель нейронной сети, которая определяет, содержит ли последовательность кадров падающего человека. Чтобы смоделировать движение видео и сделать системный сценарий независимым, используются изображения оптического потока в качестве входных данных для сетей, за которыми следует новый трехэтапный этап обучения. Кроме того, модель оценивается в трех общедоступных наборах данных, и во всех трех из них достигаются самые современные результаты.

ОСНОВНАЯ ЧАСТЬ

1. Обнаружение падений

Из-за физической слабости, связанной со старением, пожилые люди часто падают, что часто влечет за собой негативные последствия для их здоровья. Согласно Ambrose et al. [1], падения являются одной из основных причин смертности среди пожилых людей. Отчасти это можно объяснить высокой частотой падений среди взрослых старше 65 лет: каждый третий взрослый падает по крайней мере один раз в год. Кроме того, последствия этих падений являются серьезной проблемой для систем здравоохранения. Следует отметить, что падения приводят к травмам средней и тяжелой степени, страху падения, потере независимости и смерти каждого третьего человека из числа пожилых людей, пострадавших в этих несчастных случаях. Более того, затраты, связанные с этими проблемами здравоохранения, не являются незначительными: две эталонные страны, такие как Соединенные Штаты и Соединенное Королевство, с очень разными системами здравоохранения, потратили в 2008 году 23,3 и 1,6 миллиарда долларов США соответственно [2]. Принимая во внимание рост стареющего населения, ожидается, что к 2020 году эти расходы приблизятся к 55 миллиардам долларов США. Эти соображения стимулировали исследования по автоматическому обнаружению падений, чтобы обеспечить быструю и надлежащую помощь пожилым людям.

Наиболее распространенные стратегии заключаются в сочетании сенсорных и вычислительных технологий для сбора соответствующих данных и разработки алгоритмов, которые могут обнаруживать падения на основе собранных данных [3]. Эти подходы привели к появлению интеллектуальных сред для оказания помощи пожилым людям, которые традиционно ограничивались домашними условиями [4]. Однако можно полагать, что с внедрением парадигмы Интернета вещей (IoT) возможности расширения интеллектуальных сред и, более конкретно, подходов к обнаружению падений значительно возрастут.

В этой работе рассматриваются подходы, основанные на зрении, для обнаружения падений. Камеры предоставляют очень богатую информацию о людях и окружающей среде, и их присутствие становится все более важным в различных повседневных условиях из-за необходимости наблюдения.

Аэропорты, железнодорожные и автобусные вокзалы, торговые центры и даже улицы уже оборудованы камерами. Что еще более важно, камеры также установлены в центрах по уходу за пожилыми людьми. Таким образом, надежные системы обнаружения падений на основе визуального анализа могут сыграть очень важную роль в будущих системах здравоохранения и оказания помощи. Недавнее внедрение глубокого обучения изменило ландшафт компьютерного зрения, улучшив результаты, полученные во многих актуальных задачах, таких как распознавание объектов, сегментация и т.д.

В этой работе представлен новый подход в этой области, который использует преимущества сверточных нейронных сетей (CNN) для обнаружения падений. Точнее, создается CNN, которая учится обнаруживать падения по изображениям оптического потока. Учитывая небольшой размер типичных наборов данных с падениями, используются возможности сверточных нейронных сетей для последовательного обучения на разных наборах данных. Прежде всего модель обучается на наборе данных Imagenet [6], чтобы получить соответствующие функции для распознавания изображений. Затем, следуя подходу [7], CNN обучается на наборе данных UCF101 [8]. Для этой цели вычисляются изображения оптического потока последовательных кадров и используются, чтобы научить сеть обнаруживать различные действия. Наконец, применяется трансферное обучение, уменьшая веса сети и точно настраивая уровни классификации, чтобы сеть сосредоточилась на бинарной задаче обнаружения падения.

В данном исследовании есть несколько особенностей:

– Трансферное обучение применяется из области распознавания действий в область обнаружения падений. В этом смысле использование трансферного обучения имеет решающее значение для решения проблемы небольшого количества выборок в наборах данных по обнаружению падений в общественных местах.

– Использование изображений оптического потока в качестве входных данных для сети обеспечивает независимость от особенностей окружающей среды. Эти изображения представляют только движение последовательных видеокадров и игнорируют любую информацию, связанную с внешним видом, такую как цвет, яркость или контрастность. Таким образом, представляется общий подход CNN к обнаружению падения.

2. Архитектура системы

При разработке архитектуры обнаружения падений преследовались следующие цели:

1. Сделать систему независимой от особенностей окружающей среды.
2. Свести к минимуму ручную обработку изображений.
3. Сделать систему универсальной, чтобы она работала в разных сценариях.

Для решения первой задачи ключевым моментом было создание системы, которая бы учитывала движения человека и избегала какой-либо зависимости от внешнего вида изображения. В этом смысле падение видео можно выразить как несколько смежных кадров, сложенных вместе. Однако это наивный подход, поскольку корреляция между кадрами не учитывается при обработке каждого изображения отдельно. Для решения этой проблемы был использован алгоритм оптического потока [9] для описания векторов смещения между двумя кадрами. Оптический поток позволил эффективно отображать движения человека и избежать влияния статических особенностей изображения.

Чтобы свести к минимуму этапы обработки изображений вручную, используются CNN, которые, как было показано, являются очень универсальными автоматическими экстракторами признаков [5]. CNN могут изучить набор функций, которые лучше подходят для конкретной проблемы, если на этапе обучения будет предоставлено достаточно примеров. Более того, CNN также являются очень удобными инструментами для достижения общих

функций. Для этого необходимо настроить параметры сети и стратегии обучения.

Поскольку управление временем является решающим вопросом при обнаружении падения, в CNN необходимо было добавить способ управления временем и движением. С этой целью был собран набор изображений оптического потока и передан в CNN, чтобы извлечь массив признаков $F \in R^{w \times h \times s}$, где w и h — ширина и высота изображений, а s — размер стека (количество сложенных изображений оптического потока). Изображения оптического потока представляют собой движение двух последовательных кадров, которое слишком кратковременно, чтобы обнаружить падение. Однако, объединив их набор, сеть также может изучить более длинные функции, связанные со временем. Эти функции использовались в качестве входных данных классификатора — полносвязной нейронной сети (FC-NN), которая выдает сигнал «падение» или «нет падения». Полный конвейер можно увидеть на рис. 1.

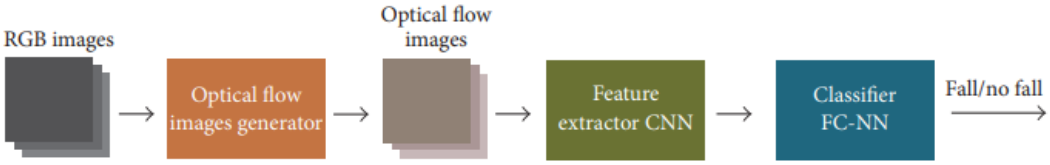

Рис.1. Архитектура системы

В результате используется трехэтапный процесс обучения для CNN на основе стека оптических потоков. Эта методология обучения принята из-за небольшого количества примеров падений, обнаруженных в общедоступных наборах данных. Кроме того, он также стремится к обобщению изученных функций для различных сценариев падения. Три этапа обучения и их обоснование подробно описаны в следующей главе.

3. Генератор изображений оптического потока

Алгоритм оптического потока представляет закономерности движения объектов в виде векторных полей смещения между двумя последовательными изображениями, которые можно рассматривать как одноканальное изображение $I \in R^{w \times h \times l}$, где w и h — ширина и высота изображения, которое представляет корреляцию между входной парой. Сложив $2L$ изображения оптического потока (т. е. L пар горизонтальных и вертикальных компонентов векторных полей, d_t^x и d_t^y соответственно), мы можем представить картину движения по составным кадрам. Это полезно для моделирования коротких событий, таких как падения. Использование изображений оптического потока также мотивировано тем фактом, что все статическое (фон) удаляется и учитывается только движение. Следовательно, входные данные инвариантны к среде, в которой будет происходить падение. Однако оптический поток также создает некоторые проблемы, например, с изменениями освещения, поскольку они могут создавать нежелательные векторы смещения. Новые алгоритмы пытаются облегчить эти проблемы, хотя не существует способа решить их во всех случаях. Однако в доступных наборах данных условия освещения стабильны, поэтому алгоритм оптического потока кажется наиболее подходящим выбором.

Первая часть конвейера — генератор изображений оптического потока — получает L последовательных изображений и применяет алгоритм оптического потока TVL-1. TVL-1 был выбран из-за его лучшей производительности при изменении условий освещения по сравнению с другими алгоритмами оптического потока. Взяв отдельно горизонтальную и вертикальную компоненты векторного поля (d_t^x и d_t^y соответственно) и сложив их вместе, создаются стеки $O \in R^{244 \times 244 \times L}$, где $O = \{d_t^x, d_t^y, d_{t+1}^x, d_{t+1}^y, \ldots, d_{t+L}^x, d_{t+L}^y\}$.

Точнее был использован программный инструмент, представленный в [36] для вычисления изображений оптического потока. Пример работы инструмента представлен на рис. 2.

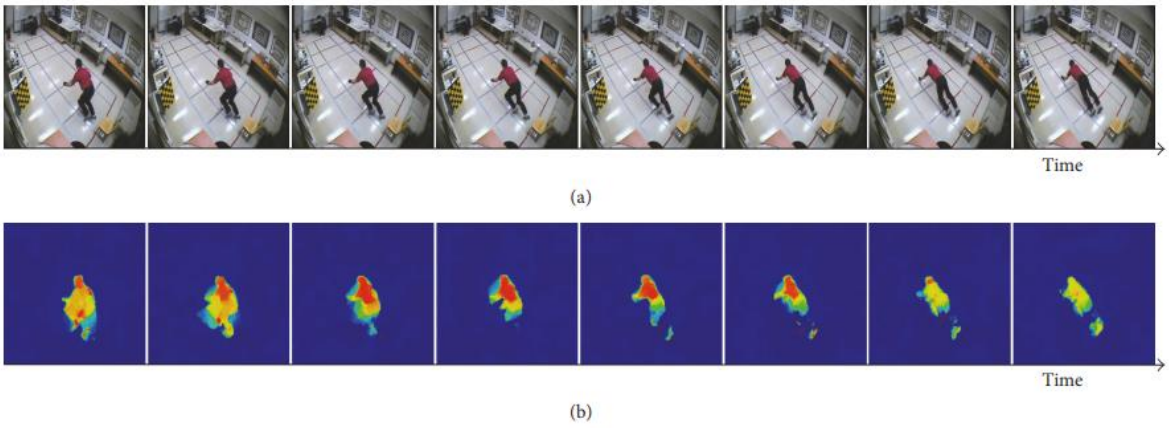

Рис. 2. Пример последовательных кадров падения из набора данных падения (a) и соответствующие им изображения горизонтального смещения оптического потока (b)

Исходные параметры вычисления оптического потока были сохранены, чтобы воспроизвести их результаты при распознавании действий. Затем была использована точно такая же конфигурация для вычисления изображений оптического потока из наборов данных обнаружения падений. Поскольку CNN изучила фильтры в соответствии с изображениями оптического потока из наборов данных распознавания действий, использование той же конфигурации для изображений обнаружения падений сводит к минимуму потерю производительности из-за трансферного обучения.

4. Архитектура нейронной сети и методология обучения

Архитектура CNN стала ключевым решением при разработке системы обнаружения падения. В последние годы было создано множество архитектурных проектов для распознавания изображений (AlexNet, VGG-16 и ResNet и другие), которые в равной степени использовались в задачах компьютерного зрения. В данной задаче была выбрана модифицированная версия сети VGG-16, следуя архитектуре временной сети [7] для распознавания действий. Использование такой архитектуры было мотивировано высокой

точностью, полученной в других смежных областях. Более конкретно, входной слой VGG-16 был изменен так, чтобы он принимал стек изображений оптического потока $O \in R^{w \times h \times 2L}$, где w и h — ширина и высота изображения, а $2L$ — размер стека. (L — настраиваемый параметр). $L = 10$, количество изображений оптического потока в стеке, как подходящее временное окно для точного захвата кратковременных событий, таких как падения. Вся архитектура была реализована с использованием фреймворка Keras [10]. На рис. 3 представлена архитектура VGG-16: сверточные слои выделены зеленым цветом, слои с максимальным объединением - оранжевым, а полностью связанные слои - фиолетовым.

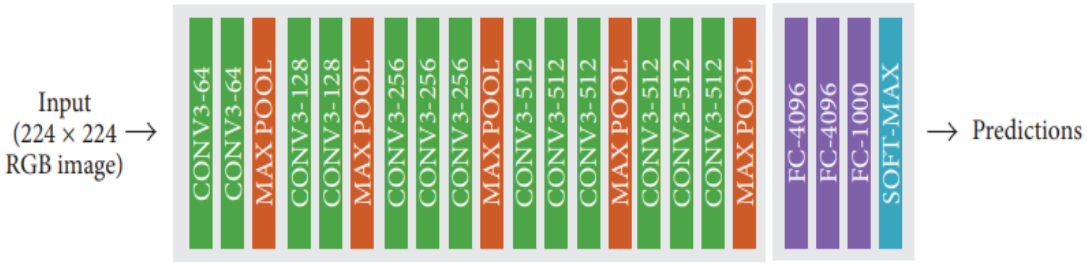

Рис. 3. Архитектура VGG-16

Процесс обучения проходил в 3 этапа. На рис. 4 кратко представлены шаги трансферного обучения.

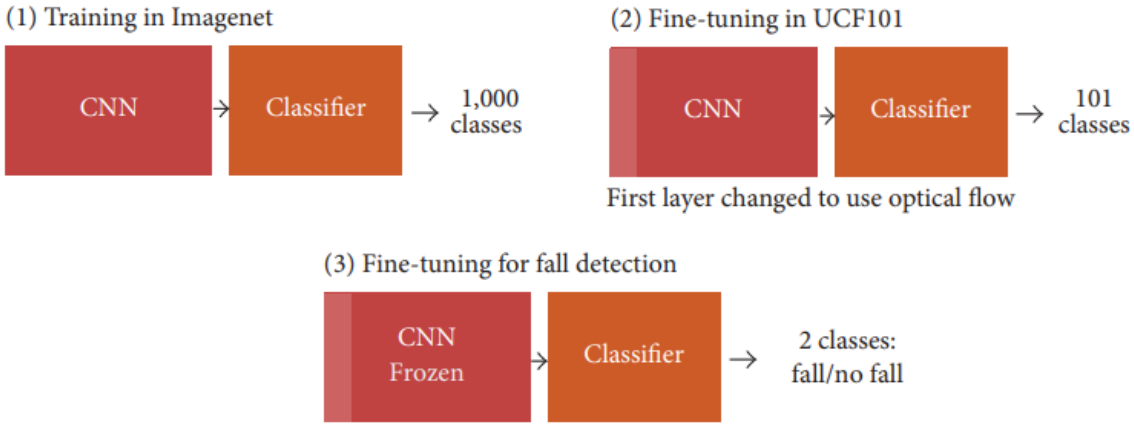

Рис. 4. Трехэтапный процесс обучения

1 этап

VGG-16 была обучена на наборе данных Imagenet [6], который содержит более 14 миллионов изображений и 1000 классов. Это стандартная практика в литературе по глубокому обучению, поскольку сеть изучает общие функции распознавания изображений. Несмотря на то, что целью является обработка изображений оптического потока, а не изображений RGB, в [7] утверждают, что общие функции внешнего вида, полученные от Imagenet, обеспечивают надежную инициализацию сети для изучения дополнительных функций, ориентированных на оптический поток. Это связано с общими функциями, изученными на первых уровнях сети, поскольку они полезны для любого домена. Затем необходимо настроить только верхнюю часть для адаптации к новому набору данных.

2 этап

На основе CNN, обученной на Imagenet, модифицируется входной слой для приема входных данных $O \in R^{244\times244\times20}$, где 224×224 — размер входных изображений архитектуры VGG-16, а 20 — размер стека, как описано в [7]. Затем сеть переобучается с помощью стеков оптических потоков из набора данных UCF101 [9]. Этот набор данных содержит 13 320 видеороликов и 101 действие человека. Этот второй шаг позволил сети изучить функции, представляющие движения человека, которые позже можно было использовать для распознавания падений.

3 этап

На последнем этапе были заморожены веса сверточных слоев, чтобы они оставались неизменными во время обучения. Чтобы ускорить процесс, были сохранены признаки, извлеченные из сверточных слоев, до первого полносвязного слоя, следовательно, имея массивы признаков $F \in$ R размером 4096 для каждого входного стека. По сути, третий шаг заключается в тонкой настройке оставшихся двух полносвязных слоев с использованием регуляризации с параметрами 0,9 и 0,8.

Для тонкой настройки с помощью набора данных падений был извлечено L кадров из последовательностей падения и «без падения» (извлеченных из исходных видео), используя скользящее окно с шагом 1 (рис. 5). Таким

образом, было получено $N-L+1$ блоков кадров, предполагая, что N — это количество кадров в данном видео и L размер блока, вместо N/L из неперекрывающегося скользящего окна. Чтобы справиться с несбалансированными наборами данных, была выполнена повторная выборка (без замены) данных, помеченных как «без падения», чтобы они соответствовали размеру данных, помеченных как падение.

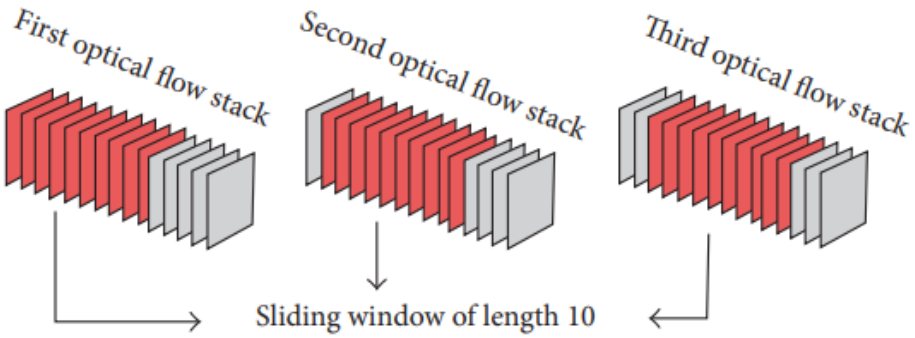

Рис. 5. Метод скользящего окна для получения стеков последовательных кадров

Даже после балансировки наборов данных изучение класса "падение" оказалось трудным. Как можно увидеть в следующих главах, сеть работала не так хорошо, как с классом “без падения”. Таким образом, понадобились альтернативные способы повысить важность класса падения в процессе обучения, такие как модификация функции потерь.

Эта функция нейронной сети выражает, насколько далеки прогнозы от реальной истины, являясь ориентиром для обновления веса во время обучения. Была выбрана бинарная кросс-энтропия:

$$loss(p,t) = -(t \cdot \log(p) + (1-t) \cdot \log(1-p))$$

где p — предсказание сети, а t — основная истина. Способ повышения важности класса состоит в добавлении к функции потерь коэффициента масштабирования или «веса класса». Следовательно, функция потерь в конечном итоге определяется следующим выражением:

$$loss(p,t) = -(w1 \cdot t \cdot \log(p) + w0 \cdot (1-t) \cdot \log(1-p))$$

где $w0$ и $w1$ — соответственно веса для «класса падения» и «класса отсутствия падения», p — это предсказание сети, а t — основная истина. Вес класса 1,0 означает отсутствие изменений в весовом значении этого класса. Использование более высокого веса для класса 0, то есть $w0$, штрафует функцию потерь за каждую ошибку, допущенную в этом классе, больше, чем за ошибки в классе 1. Нейронная сеть всегда пытается минимизировать потери, адаптируя свои веса: это основа алгоритма обратного распространения ошибки [44]. Следовательно, используя эту модифицированную функцию потерь, сеть уделяет приоритетное внимание изучению одного из классов. Однако из-за этого может произойти ухудшение обучения другого класса. По этой причине в главе с результатами представлены метрики, которые показывают производительность каждого класса отдельно.

Хотя использование $w0$ больше 1,0 смещает обучение в сторону падений (в случае $w1$ = 1), можно утверждать, что это смещение удобно при обнаружении падения из-за важности обнаружения падения даже ценой некоторых ложных тревог. Пропущенное обнаружение будет иметь решающее значение для здоровья пожилых людей, и поэтому его следует избегать.

5. Наборы данных

Было выбрано три набора данных, которые часто используются в литературе, что делает их подходящими для целей сравнительного анализа:

– UR Fall Dataset (URFD). URFD содержит 30 видеороликов о падениях и 40 видеороликов повседневной жизни (ADL), которые обозначены как "без падений".

– Multiple Cameras Fall Dataset (Multicam). Multicam содержит 24 сцены (22 - как минимум с одним падением, а остальные две - только со сбивающими с толку событиями). Каждое сцена была записана с 8 разных ракурсов. Для всех

видеороликов используется одна и та же сцена с некоторой перестановкой мебели.

– Fall Detection Dataset (FDD). FDD содержит 4 различных видео (в отличие от предыдущих) с несколькими участниками.

Доступные наборы данных записываются в контролируемых средах с различными ограничениями:

– В видео только один актер.

– Изображения записываются при подходящих и стабильных условиях освещения, избегая темноты или резких изменений освещения.

Падения появляются в разных местах сцены, вдали и рядом с камерой. В частности, в наборе данных Multicam, где доступно восемь камер, расстояние до камеры значительно варьируется. В датасете FDD некоторые падения также находятся далеко от камеры, хотя это не общий случай, как в Multicam. Все видео были сегментированы по кадрам и разделены на «с падением» и «без падений» в соответствии с предоставленными аннотациями.

Комбинация трех наборов данных также является хорошим индикатором общности подхода.

6. Результаты исследования

Чтобы проверить разработанную систему обнаружения падений, был проведен ряд экспериментов по анализу конфигурации сети с целью поиска наиболее подходящей конфигурации для решения задачи, используя наборы данных из предыдущей главы.

Методика оценки

С точки зрения обучения с учителем, обнаружение падения можно рассматривать как задачу двоичной классификации, в которой классификатор должен решить, представляют ли определенные последовательности видеокадров падение или нет. Наиболее распространенными показателями для оценки эффективности такого классификатора являются чувствительность, также известная как уровень отзыва или истинно положительный результат, и

специфичность или истинно отрицательный уровень. Эти метрики не подвержены смещению из-за несбалансированного распределения классов, что делает их более подходящими для наборов данных обнаружения падения, где количество выборок с падением обычно намного меньше, чем количество выборок без падения. Для обнаружения падений чувствительность является мерой того, насколько хорошо система прогнозирует падения, тогда как специфичность измеряет производительность при отсутствии падений. Таким образом, две метрики оценки, которые использовались, определяются следующим образом:

$$Sensitivity\ (Recall) = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

где TP относится к истинно положительным результатам, TN — к истинно отрицательным, FP — к ложноположительным, а FN — к ложноотрицательным. С помощью видео мы оценена производительность системы для каждого стека оптических потоков; то есть проверен прогноз для стека относительно его реальной метки. Для создания стека использовался размер блока в 10 последовательных кадров. Это число было найдено эмпирически в [11]. Следовательно, упомянутые выше значения определяются следующим образом:

– TP: стек оптических потоков, помеченный как падение и предсказанный как падение

– FP: стек оптических потоков, помеченный как «без падения» и прогнозируемый как падение

– TN: стек оптического потока, помеченный как «без падения» и прогнозируемый как «без падения»

– FN: стек оптического потока, помеченный как падение и прогнозируемый как «без падения».

Результаты поиска оптимальных конфигураций

В поисках лучшей конфигурации сети каждый набор данных был разделен на обучающий и тестовый набор с соотношением 80:20 и сбалансированным распределением меток (0 или 1, падение или «без падения»). Поиск в пространстве конфигураций включал в себя множество экспериментов с разной скоростью обучения.

Также варьировались значения размера батча (с возможностью пакетного обучения), сначала применяя базовый вес класса $w0$, равный 2, а затем изменяя его в тех направлениях, которые казались более выигрышными. Опции ReLU с пакетной нормализацией и ELU использовались одинаково, хотя выбирался наиболее обнадеживающий вариант, если результаты были для него благоприятными. Лучшие результаты работы системы для трех наборов данных (URFD, FDD и Multicam) с использованием разных настроек представлены в таблице 1 (в столбце размера батча «full» означает пакетное обучение).

Таблица 1. Результаты работы системы для трех наборов данных

Набор данных	**Скорость обучения**	**Размер батча**	$w0$	**Функция активации**	**Sens.**	**Spec.**
URFD	10-5	64	1	ReLU	100.0%	94.86%
Multicam	10-3	full	1	ReLU	98.07%	96.20%
FDD	10-4	1,024	2	ELU	93.47%	97.23%

ЗАКЛЮЧЕНИЕ

В ходе выполнения данной работы были получены следующие результаты:

1. Представлена архитектура сверточной нейронной сети для задачи обнаружения падений на основе оптических потоков и трансферного обучения.
2. Проведено трехэтапное обучение полученной модели сверточной нейронной сети на наборе данных Imagenet и затем на наборе данных UCF101.
3. Проведен эксперимент по поиску оптимальных конфигураций сети для трех общедоступных наборах данных обнаружения падений, а именно URFD, Multicam и FDD.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1) A. F. Ambrose, G. Paul, and J. M. Hausdorff, “Risk factors for falls among older adults: A review of the literature,” Maturitas, vol. 75, no. 1, pp. 51–61, 2013.

2) J. C. Davis, M. C. Robertson, M. C. Ashe, T. Liu-Ambrose, K. M. Khan, and C. A. Marra, “International comparison of cost of falls in older adults living in the community: A systematic review,” Osteoporosis International, vol. 21, no. 8, pp. 1295–1306, 2010.

3) M. Mubashir, L. Shao, and L. Seed, “A survey on fall detection: principles and approaches,” Neurocomputing, vol. 100, pp. 144– 152, 2013.

4) L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, “Sensorbased activity recognition,” IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 42, no. 6, pp. 790–808, 2012.

5) Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, 2015.

6) J. Deng, W. Dong, and R. Socher, “ImageNet: a large-scale hierarchical image database,” in Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 248–255, Miami, Fla, USA, June 2009.

7) L. Wang and et al., Towards good practices for very deep twostream convnets, 2015.

8) S. Khurram, A. R. Zamir, and M. Shah, Amir Roshan Zamir, and Mubarak Shah., UCF101: A dataset of 101 human actions classes from videos in the wild, 2012.

9) S. S. Beauchemin and J. L. Barron, “The Computation of Optical Flow,”ACM Computing Surveys, vol. 27, no. 3, pp. 433–466, 1995

10) Keras, [Электронный ресурс]. – URL: https://keras.io, Дата обращения: 20.12.2023.

11) K. Simonyan and A. Zisserman, “Two-stream convolutional networks for action recognition in videos,” in Proceedings of the 28th Annual Conference on Neural Information Processing Systems 2014, NIPS 2014, pp. 568–576, can, December 2014.