

**РЕФЕРАТ**

Отчет 98 с., 1 рис., 3 табл., 99 источн., 2 прил.

ФИЗИЧЕСКИЙ БУЛЛИНГ, МАШИННОЕ ОБУЧЕНИЕ, КОМПЬЮТЕРНОЕ ЗРЕНИЕ, РАСПОЗНАВАНИЕ ОБРАЗОВ, НАСИЛИЕ В ШКОЛАХ, НАСИЛИЕ В ОБРАЗОВАТЕЛЬНЫХ УЧРЕЖДЕНИЯХ, ВЫЯВЛЕНИЕ АГРЕССИВНОГО ПОВЕДЕНИЯ, АГРЕССИВНОЕ ПОВЕДЕНИЕ

Объект исследования – Видео записи, собранные из школьных камер наблюдения.

Цель проекта – Разработка решений на основе искусственного интеллекта (ИИ) и создание на их основе прототипа программно-технического комплекса, способного в автоматическом режиме выявлять факты агрессивного поведения и возможного физического буллинга в образовательных учреждениях Республики Казахстан.

Методы исследования – анализ видеопотоков посредством ИИ позволит раньше выявлять факты агрессивного поведения. Раннее выявление подобных фактов в свою очередь облегчит работу школьных психологов в части раннего предупреждения буллинга. В предлагаемом исследовании программные модели искусственного интеллекта будут обучаться на основе архитектуры LGD-3D, Two-stream, Two-stream I3D. Для классификации видео будет использоваться метод, основанный на нейронной сети с глубокими сверточными графами (DCGN). По результатам исследования данный метод превосходит альтернативные, такие как LTSM и GRU.

Полученные результаты и новизна: Обзор научных исследований для обнаружения физического буллинга, научная статья по результатам обзора, Обзор методов машинного обучения для обнаружения физической агрессии в видео, требования для сбора данных.

Научная значимость результатов проекта заключается в разработке моделей, для автоматического обнаружения возможной физической агрессии в видео, методов формирования оптимального набора признаков для разработки моделей машинного обучения для выявления агрессора и жертвы буллинга, разработке необходимого набора видеоданных для данной задачи, методов обнаружения и распознавания лиц школьников участвующих в агрессию.

Область применения: Целевые потребители полученных результатов – фундаментальные результаты могут быть использованы мировым научным сообществом; прикладные результаты могут быть использованы, образовательными учреждениями в том числе школ, колледжи, университеты и в тех областях жизнедеятельности, где необходим анализ видео.

**ТҰЖЫРЫМДАМА**

Есеп 98 б., 1 сурет, 3 кесте, 99 әдебиет, 2 қосымша.

ФИЗИКАЛЫҚ ҚОРҚЫТУ, МАШИНАЛЫҚ ОҚЫТУ, КОМПЬЮТЕРЛІК КӨРУ, ҮЛГІНІ ТАНУ, МЕКТЕПТЕГІ ЗОРЛЫҚ-ЗОМБЫЛЫҚ, БІЛІМ БЕРУ МЕКЕМЕЛЕРІНДЕГІ ЗОРЛЫҚ-ЗОМБЫЛЫҚ, АГРЕССИВТІ МІНЕЗ-ҚҰЛЫҚТЫ АНЫҚТАУ, АГРЕССИВТІ МІНЕЗ-ҚҰЛЫҚ.

Зерттеу объектісі – Мектеп камераларынан жиналған бейне деректер.

Жоба мақсаты – Жасанды интеллект (ЖИ) негізінде шешімдер әзірлеу және олардың негізінде Қазақстан Республикасының Білім беру мекемелерінде агрессиялық мінез-құлық және ықтимал физикалық буллинг фактілерін автоматты режимде анықтауға қабілетті бағдарламалық-техникалық кешеннің прототипін жасау.

Зерттеу әдісі – ЖИ әдістерімен бейнені талдау агрессивті мінез-құлық фактілерін ертерек анықтауға мүмкіндік береді. Мұндай фактілерді ертерек анықтау, өз кезегінде, мектеп психологтарының қорқытуды ерте ескерту бөлігіндегі жұмысын жеңілдетеді. Бейнені жіктеу үшін терең конвульсиялық графигі бар нейрондық желіге негізделген әдіс қолданылады (DCGN). Зерттеу нәтижелері бойынша бұл әдіс LTS және GRU сияқты балама әдістерден асып түседі.

Алынған нәтижелер мен жаңалықтар: Физикалық қорқытуды анықтауға арналған ғылыми зерттеулерге шолу, шолу нәтижелері бойынша ғылыми мақала, бейнедегі физикалық агрессияны анықтауға арналған машиналық оқыту әдістеріне шолу, деректерді жинауға қойылатын талаптар.

Жоба нәтижелерінің ғылыми маңыздылығы: Бейнедегі физикалық агрессияны автоматты түрде анықтау үшін модельдерді әзірлеу, агрессорды және қорқытуды құрбанды анықтау үшін машиналық оқыту модельдерін жасау үшін оңтайлы белгілер жиынтығын қалыптастыру әдістері, осы тапсырма үшін қажетті бейне мәліметтер жиынтығын жасау, агрессияға қатысатын мектеп оқушыларының бет-әлпетін анықтау және тану әдістері.

Қолданылу саласы: алынған нәтижелердің нысаналы тұтынушылары – іргелі нәтижелерді әлемдік ғылыми қоғамдастық пайдалана алады; қолданбалы нәтижелерді білім беру мекемелері, оның ішінде мектептер, колледждер, университеттер және бейне талдау қажет болатын тіршілік ету салаларында пайдалана алады.

**СОДЕРЖАНИЕ**

[ВВЕДЕНИЕ 6](#_Toc58413481)

[ОСНОВНАЯ ЧАСТЬ ОТЧЕТА О НИР 8](#_Toc58413482)

[1 Буллинг: общая информация 8](#_Toc58413483)

[1.1 Классификация проявлений буллинга 8](#_Toc58413484)

[1.2 Характеристика жертвы и агрессора 9](#_Toc58413485)

[1.3 Буллинг и нейронаука 10](#_Toc58413486)

[1.4 Влияние буллинга на жертву и агрессора 11](#_Toc58413487)

[2 Проведение обзорных исследований по разработке программно-технического комплекса способного в автоматическом режиме выявлять факты агрессивного поведения и возможного физического буллинга 14](#_Toc58413488)

[2.1 Анализ литературы по существующим методам автоматического обнаружения фактов физического буллинга 14](#_Toc58413489)

[2.1.1 Классификация методов выявления насилия (МВН) 14](#_Toc58413490)

[2.1.2 Обнаружение насилия с использованием методов машинного обучения 15](#_Toc58413491)

[2.1.3 Методы обнаружения насилия с использованием SVM 21](#_Toc58413492)

[2.2 Обзор существующих программных решений по распознаванию и классификации действий на видео 33](#_Toc58413493)

[3 Разработка требований для сбора данных 36](#_Toc58413494)

[3.1 Разработка требований к процессу сбора данных 36](#_Toc58413495)

[3.2 Разработка требований к качеству данных и их объему 36](#_Toc58413496)

[3.3 Разработка требований к разметке собранных данных 37](#_Toc58413497)

[ЗАКЛЮЧЕНИЕ 38](#_Toc58413498)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 39](#_Toc58413499)

[ПРИЛОЖЕНИЕ А Календарный план 48](#_Toc58413500)

[ПРИЛОЖЕНИЕ Б Обзорная статья 53](#_Toc58413501)

ВВЕДЕНИЕ

Во всем мире растет число школьных программ по борьбе с хулиганством [1-2]; многие из них были разработаны на основе вмешательства, разработанного в работе [3]. Однако эти программы имели разную степень успеха, от уменьшения до обострения проблемы [4]. Программы могут потерпеть неудачу из-за недостаточной точности развертывания инициативы, ограниченной поддержки учителей и персонала и/или отсутствия долгосрочного наблюдения.

Olweus Bullying Prevention Program (OBPS – Программа Ольвеуса для Профилактики Издевательств) была первой общешкольной программой по борьбе с хулиганством, разработанной в Норвегии [5]. Программа Ольвеуса включала в себя информацию о хулиганстве для учителей и родителей, записанное видео о школьных хулиганствах и анкету о жертвах хулиганства для учащихся. С момента завершения первоначального проекта OBPS разрабатывался и оценивался неоднократно [6]. Полученные результаты показали неизменно положительные эффекты [7].

К числу других примечательных программ относятся программа Be-Prox в Швейцарии [8], программа KiVa в Финляндии [9] и программа Kia Kaha в Новой Зеландии [10]. В Ирландии были введены такие программы, как Stay Safe [11]. Результаты оценочных исследований подчеркивают важное влияние этих программ на эмпатию [12-13]. Кроме того, принятие целостного школьного подхода рассматривается как важнейший компонент в работе [14]. Брендген и Пулин [15] отметили, что программы профилактики издевательств выиграют от дополнительных вмешательств, специально направленных на развитие социальных навыков и создание поддерживающей сети.

Методы обнаружения насилия с помощью компьютерного зрения анализируют видео с камер наблюдения. В течение последних нескольких лет эти камеры и другое оборудование наблюдения устанавливаются в различных местах для обеспечения общественной безопасности, например в учебных заведениях, больницах, банках, на рынках, улицах, чтобы следить за активностью людей [16]. Мониторинг включает в себя анализ поведения людей, независимо от того, являются ли их действия подозрительными или нормальными. Обнаружение подозрительной активности в течение 24/7 или поиск такой активности в огромных данных, состоящих из записанных видео, является очень сложной задачей [17-20]. Для ее решения были разработаны различные методы распознавания человеческой деятельности в реальной жизни. Эти методы помогают обнаружить подозрительные действия на видеозаписях наблюдения.

Обнаружение насилия по видеозаписям [21] является одним из видов обнаружения активности. Разработаны определенные методики и приемы для обнаружения агрессивного поведения и других видов нежелательной активности на видео [22-24]. В них предлагаются различные подходы, которые работают с различными входными параметрами. Параметры — это в основном различные атрибуты или особенности видео, такие как ускорение, поток, время, внешний вид и т.д. В процессе обнаружения активности первым шагом является разделение всего видео на сегменты и кадры [25]. Во-вторых, из видеокадров определяется объект. В-третьих, в соответствии с применяемым методом извлекаются особенности видео. Наконец, по кадрам обнаруживается аномальная активность. Шаги варьируются в зависимости от используемых методов.

ОСНОВНАЯ ЧАСТЬ ОТЧЕТА О НИР

1 Буллинг: общая информация

1.1 Классификация проявлений буллинга

Согласно определению Олвеуса [26], буллинг в школе - это поведение, при котором жертва буллинга многократно и в течение времени подвергается негативным действиям со стороны одного или нескольких школьников. Негативные действия могут быть в виде физического контакта, слов или других способов, например, изменения выражения лица или непристойные жесты, а также стремление преднамеренно исключать жертву буллинга из группы, сообщества. В целом буллинг характеризуется тремя критериями [50]:

* агрессивное поведение или преднамеренное «причинение вреда»;
* которое происходит «многократно и в течение времени»;
* имеется дисбаланс власти.

Буллинг подразумевает под собой не только физическую агрессию, но и психологическое давление, запугивание, распространение сплетен, вымогательство, издевательства.

Буллинг делится на прямые и непрямые формы агрессии. Физическая и вербальная формы агрессии являются прямыми формами буллинга, состоящими из явного выражения власти. Физический буллинг включает в себя любой тип физического нападения, например, удары, толчки, удары ногами, удушение и насильственное отнятие чего-либо у жертвы. Устные издевательства включают словесные домогательства или запугивания в форме обзывания, угроз, насмешек, злонамеренных насмешек и психологическое запугивание с использованием слов для причинения вреда жертвам. Воровство, вандализм, выражение лица или непристойные жесты также являются способами, с помощью которых дети подвергаются издевательствам. [51]

Агрессивное поведение определяется как направленное вредящее действие на другого человека, воздействуя психологически или физически (Berkowitz, 1993).

С 2005 года исследования издевательств в школе в Гонконге включали агрессивное поведение в более общем плане. В этих исследованиях агрессивное поведение, связанное со школьными хулиганами, классифицируется на два конкретных подтипа в соответствии с функциями и основными целями: реактивные и проактивные агрессоры. Различия между реактивной и проактивной агрессией были продемонстрированы в исследованиях последних трех десятилетий. Додж и Кои различали типы агрессивного поведения в зависимости от их различных целей. Реактивная агрессия - это импульсивная и эмоциональная реакция на воспринимаемое разочарование, угрозу или провокацию. Враждебность и гнев действуют как механизмы защиты или возмездия. Напротив, проактивная агрессия - это инструментальная агрессия. Подростки, проявляющие активную агрессию, считаются настоящими хулиганами. Они применяют принудительное поведение, чтобы показать свое превосходство и доминирование или достичь личных целей; эти действия часто совершаются с умышленной целью, но без немедленной провокации или подстрекательства. [27]

1.2 Характеристика жертвы и агрессора

В литературе широко распространены пять основных ролей участников издевательств: хулиган, жертва, помощник, защитник и посторонний. Исторически сложилось так, что исследования были сосредоточены на двух ролях: хулиган (который совершает агрессию) и жертва (кто подвергается издевательству). Текущие исследования были расширены для изучения буллинга как группового процесса, который включает в себя другие роли, такие как помощник, который помогает хулигану; защитник, который встает на защиту жертвы; и посторонний, который избегает буллинг (Pouwels, Lansu, & Cillessen, 2016; Salmivalli, 2010; Salmivalli et al., 1996). Эти роли не исключают друг друга, а это означает, что человек может участвовать в более чем одной роли в зависимости от ситуации, например, в качестве жертвы хулигана (Pouwels et al., 2016). [28]

Агрессоры используют свою силу и физическое превосходство, чтобы доминировать и достигать своих целей. Sutton (2001) and Sutton, Smith and Swettenham (1999) выдвинули гипотезу о том, что хулиганы, успешно манипулируя другими, имеют более высокую самооценку. Однако это лишь отражение их завышенного, нарциссического представления о себе (Salmivalli, 2001). [29] В случаях издевательств или запугивания, когда роли участников буллинга связаны с их самооценкой, было замечено, что агрессорами чаще выступают мужчины. [30] Агрессоры чаще других пропускают занятия, склонны к девиантному поведению, раннему вступлению в половые контакты. Большинство агрессоров сообщают о чувствах одиночества, эпизодах насилия и высокой распространенности физического насилия в семьях. [31] “Слишком мало любви и заботы, и слишком много свободы в детстве” со стороны родителей увеличивают вероятность развития модели агрессивности (Olweus, 1978).

Пол играет значительную роль в проявлении агрессии. По некоторым данным агрессоры мужского пола чаще применяют физическую и вербальную виды агрессий, агрессоры женского пола склонны наоборот проявлять агрессию по отношению к себе, что может иметь значение для предотвращения самоубийств среди молодежи. [32]

Согласно всеобщему представлению, жертвы - это физически слабые, тихие и замкнутые люди, часто беспомощные, подчиняющиеся. Хотя Olweus (1993) не обнаружил какого-либо физического фактора риска для того, чтобы стать жертвой (кроме слабости), есть свидетельства того, что жертвы часто имеют физические отклонения (например, ожирение) [33] или физические недостатки (например, дефекты зрения или речи). [29] В исследовании Fatma A. и соав., большинство детей с ожирением (65%) и детей с избыточным весом (70%) подвергались издевательством и сообщали о самом низком уровне самооценки. [34] Проведенное в Ирландии исследование 910 детей в возрасте 9 лет еще раз показало, что дети у которых индекс массы тела (ИМТ) превышал нормальные показатели чаще подвергались буллингу, однако была выявлена связь между представлением ребенка себя, как худого или очень худого с позицией жертвы. [35] В последующем Yen et al. доказали, что лица с проблемами психического здоровья, такими как фобия, депрессия, невнимательность, чаще становятся жертвами пассивного или активного издевательства. Также лица, совершившие пассивное и активное издевательство имели серьезные симптомы депрессии и гиперактивности/импульсивности. [36]

1.3 Буллинг и нейронаука

Было предположено, что виктимизация сверстников может быть связана с измененной нейробиологией, с помощью которой может влиять на психологическое функционирование человека в целом (Rudolph et al., 2016), особенно с вовлечением вентролатеральной префронтальной коры (vlPFC). Например, функциональные нейровизуализационные исследования показывают, что социальное исключение связано с активацией vlPFC, а активация vlPFC отрицательно коррелирует с социальным дистрессом во время инцидента. [37]

Неритмичные процессы секреции кортизола в течение дня были описаны Fries et al. в 2009, а также реакция на стресс связанная с физическими проблемами (Miller et al., 2007) и негативными последствиями, такими как проблемное поведение (например, McBurnett et al. al., 2000; Alink et al., 2008) и психопатологии (Buitelaar, 2013). Гипоталамо-гипофизарно-надпочечниковая система (HPA) может также влиять на различия в защитных механизмах в преодолении дистресса, изменения которого приходятся на подростковый возраст (Lupien et al., 2009). Есть некоторые свидетельства того, что уровень кортизола может смягчать влияние виктимизации со стороны сверстников на исходы. Например, было показано, что виктимизация связана с усилением депрессивных симптомов только у лиц с повышенным базальным уровнем кортизола (Brendgen et al., 2017) и повышенной реакцией кортизола на внешний стимул (Rudolph et al., 2011). Эти результаты показывают, что люди с гиперчувствительной системой реагирования на стресс могут быть менее способны эффективно справляться с виктимизацией со стороны сверстников. В то время как аналогичные исследования не проводились в отношении нейробиологических результатов, индивидуальные различия в функции HPA, вероятно, влияют на структуру головного мозга. [38]

Анализ с проведением функциональной МРТ головного мозга также показал, что повышенная активность ростральной передней поясной коры головного мозга, вызывающая испуганные лица, была связана с меньшим количеством издевательств. Эти результаты предполагают, что издевательства и виктимизация связаны с различными моделями нейронной активности при эмоциях испуга и гнева, что может помочь в понимании того, как модели обработки социальной информации предсказывают эти переживания.

Считается, что эти паттерны мозговой активности характеризуют разные пути к антисоциальному поведению, но также возможно, что они могут взаимодействовать. Например, подросток с повышенной активностью миндалины при выражении гнева и пониженной активностью миндалины при испуге может подвергаться более высокому риску агрессивного поведения, чем подросток, демонстрирующий только один из этих паттернов активности, хотя, насколько нам известно, этот тип взаимодействия не применялся. [39]

Более того, мы наблюдали статистически значимую связь между атрофией гиппокампа и дисфункциональными условиями рабочей среды, где по по анкетированию респондентов присутствовали травля, буллинг и унижения.

Таким образом, данные результатов нейробиологических исследований пытаются объяснить разнородную и сложную последовательность природы буллинга как разновидность асоциального поведения человека.[40]

1.4 Влияние буллинга на жертву и агрессора

Последствия издевательств могут быть серьезными как и для жертв, так и для хулиганов. Было выяснено, что виктимизация способствует развитию депрессии, соматических болезней, неудовлетворенности жизнью и низкой частоте поступления жертв буллинга в колледж через 3 года после инцидента. Например, частота проблем со здоровьем увеличивается на 75% из-за издевательств, а неприязнь к школе увеличивается на половину стандартного отклонения для тех, кто позиционирует себя как “жертва”. [41] Виктимизация со стороны сверстников считается фактором высокого риска для развития у школьников шизотипических черт личности и шизотипического расстройства личности по данным исследования, проведенного в Гонконге. [42]

В исследовании Husky и соав. с выборкой 5183 учеников начальных школ семи европейских стран (Болгария, Германия, Италия, Нидерланды, Румыния, Турция), дети, идентифицированные как жертвы с высокой вероятностью имели проблемы с психическим здоровьем. Хулиганы-жертвы с большей вероятностью имели фобии, тревожные и депрессивные расстройства, в то время как статус жертвы был связан с ГТР (генерализованным тревожным расстройством), а статус хулигана с тревогой разлуки. [43] В исследовании, проведенном в Великобритании в 2013 году, изучалось влияние пережитых в начальной школе издевательств на возникновение психотических переживаний в позднем подростковом возрасте, где симптомы симптомы психоза или депрессии в возрасте 18 лет были лишь частично опосредованы виктимизацией в детстве. [44]

Связь между участием в издевательствах, будь то в качестве жертвы или агрессора, и развитием психотической симптоматики была обнаружена и в другом 35-летнем когортном исследовании (p <0,0001). В ходе статистических расчетов у лиц с наивысшим уровнем травли виктимизации частота психотических симптомов была в 1,21 (95% доверительный интервал 0,73–1,99) раза выше, чем у тех, кто не подвергался виктимизации. [45]

Предыдущие результаты подтверждает исследование Avon «Родители и дети» в Великобритании (ALSPAC) и Great Smoky Mountains в США (GSMS), где те, кто подвергался жестокому обращению и издевательствам, имели более высокий риск проблем психического здоровья, беспокойства и депрессии по обеим когортам и членовредительства по когорте ALSPAC со сравнительной группой детей. Дети, над которыми издевались только сверстники, чаще, чем дети, подвергшиеся жестокому обращению, имели проблемы с психическим здоровьем в обеих когортах (ALSPAC OR 1,6, 95% ДИ 1 · 1–2 · 2; p = 0,005; GSMS 3 · 8, 1 · 8–7 · 9, p <0 · 0001), с различиями в тревожности (GSMS OR 4 · 9; 95% CI 2 · 0–12 · 0), депрессии (ALSPAC 1 · 7, 1 · 1–2 · 7) и членовредительства (ALSPAC 1 · 7, 1 · 1–2 · 6) между двумя когортами. [46]

В Канадском исследовании изучалось связь суицидальности и издевательств и сексуальных домогательств среди подростков. Повторяющиеся издевательства онлайн и оффлайн, повлекшие за собой симптомы посттравматического стрессового расстройства и недостаточное присутствие материнской поддержки способствовали активным суицидальным мыслям у девочек-подростков от 12 до 18 лет, в то время как сексуальные домогательства не были связаны с суицидальностью. [47]

Аналогично связь между женским полом и суицидальными идеями была подтверждена в исследовании Китая с выборкой 23 392 школьников, где изучалось отношение между агрессивным поведением с суицидальными идеями и попытками, а также влияние пола. Жертвы, хулиганы и жертвы хулиганов имели повышенный риск суицидальных мыслей, чем нейтральные; аналогичные ассоциации можно найти в ассоциациях между агрессивным поведением и попытками самоубийства. Дальнейший стратификационный анализ показал, что ассоциации травли или запугивания других с суицидальными идеями и попытками самоубийства были немного сильнее у девочек, чем у мальчиков. [48]

2 Проведение обзорных исследований по разработке программно-технического комплекса способного в автоматическом режиме выявлять факты агрессивного поведения и возможного физического буллинга

2.1 Анализ литературы по существующим методам автоматического обнаружения фактов физического буллинга

2.1.1 Классификация методов выявления насилия (МВН)

Быстрый рост объема видеоданных привел к возрастанию потребности в наблюдении и обнаружении аномалий. Аномальные события редко происходят по сравнению с нормальной деятельностью. Разработка автоматизированных систем обнаружения аномалий стала насущной потребностью для уменьшения потерь труда и времени. Обнаружение аномалий в видео является сложной задачей, поскольку определение аномалии может быть неоднозначным и неопределенным.

Одним из важных аспектов обнаружения аномалий является распознавание и обнаружение насилия. Методы моделирования, используемые для обнаружения аномалий или насильственных действий, могут быть широко классифицированы как неглубокие и глубокие.

Неглубокие методы моделирования – это методы, которые не способны самостоятельно извлекать объекты. Объекты, извлеченные с помощью ручных методов, должны быть предоставлены неглубокой модели для классификации. В качестве неглубокой модели могут использоваться такие классификаторы, как машина опорных векторов (SVM), искусственная нейронная сеть (ANN) с одним скрытым слоем и т.д. Эти модели лучше всего подходят для обучения с подкреплением, которое тебует хороших размеченных данных. Основным недостатком таких методов является то, что они не адаптируются к изменениям паттерна автоматически. Кроме того, процесс разметки может происходить вручную. Ллойд и др. [52] предложили дескриптор реального времени, который моделирует динамику толпы для обнаружения аномалий путем кодирования изменений в текстуре толпы с использованием временных сводок признаков матрицы встречаемости серого уровня, в которых k-кратная перекрестная валидация была выполнена для обучения классификатора случайного леса. Аналогично, Билински и Бремонд [53] использовали расширение улучшенных векторов Фишера (IFV), которое позволяет представлять видео с использованием как локальных признаков, так и их пространственно-временных позиций для распознавания и обнаружения насилия. Их результаты показали значительное улучшение в четырех общедоступных стандартных эталонных наборах данных.

В отличие от неглубоких моделей, большинство глубоких моделей не требуют отдельного экстрактора признаков, так как они основаны на методе изучения признаков, который заключается в том, что они изучают свои собственные признаки из заданных данных и классифицируют их на основе них. Кроме того, помимо сквозного обучения, вышеупомянутые извлеченные функции могут быть даны в качестве входных данных для SVM и других классификаторов неглубоких моделей. Другой способ реализации глубоких моделей – это использование признаков из созданных вручную дескрипторов признаков и предоставление их глубокому классификатору. Эти модели работают как с контролируемыми, так и с неконтролируемыми методами обучения, но лучше подходят для последних. Несмотря на то, что они работают с неразмеченными данными, они требуют больших объемов данных и вычислительной мощности. В работе [54] предлагается сверточный пространственно-временной автоэнкодер для изучения регулярных паттернов в обучающих видеороликах для обнаружения аномалий. Несмотря на то, что модель может обнаруживать аномальные события и устойчива к шуму, в зависимости сложности деятельности, может возникнуть больше ложных результатов.

Методы обнаружения насилия (МОН) классифицируются на три категории в зависимости от используемого классификатора: МОН с использованием машинного обучения, МОН с использованием SVM и МОН с использованием глубокого обучения. SVM и глубокое обучение классифицируются отдельно, так как эти алгоритмы широко используются в компьютерном зрении. Подробная информация о каждом методе представлена в таблице 1. Методы представлены в хронологическом порядке. Также представлены методика обнаружения объектов и метод извлечения признаков. Общая система обнаружения аномалий или насильственных действий показана на рисунке 1. Далее, в обзоре рассматриваются методы для автоматического обнаружения насилия, физического буллинга и агрессии.

2.1.2 Обнаружение насилия с использованием методов машинного обучения

Распознавание насильственных действий в видеозаписях наблюдения с помощью компьютерного зрения становится активной темой в области обнаружения действий [55]. В данном разделе описаны методы обнаружения насилия, использующие в качестве классификатора различные традиционные алгоритмы машинного обучения, такие как KNN, Adaboost и другие.

В таблице 1 представлен перечень методик обнаружения, использующих различные методы классификации. Представлены также положения метода классификации, используемые в предлагаемой методике.

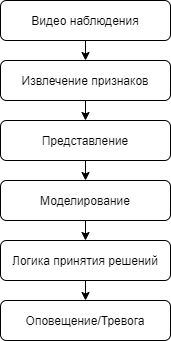


Рисунок 1 – Общая схема системы обнаружения насилия

Таблица 1 - Методы выявления насилия с использованием различных методов классификации

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Метод | Метод обнаружения объекта | Метод извлечения признаков | Метод классификации | Тип проис-шествия | Точность % |
| 1 | 2 | 3 | 4 | 5 | 6 |
| Движения больших двоичных объектов (AMV) ускорение векторная физическая величина, характеризующая способ для обнаружения быстрых изменений в видео [56] | Эллипсовый метод обнаружения насилия | Алгоритм нахождения ускорения | Использование пространственно-временных признаков для классификации | Любой | 90% |
| Метод RIMOC фокусируется на скорости и направлении движения объекта на основе HOF (гистограмма оптического потока) [57] | Метод ковариационной матрицы на основе пространственно-временной векторный метод | Пространственно-временной векторный метод | SVM с помощью обучения с учителем | И многолюдно, и менее многолюдно | 82 - 90 |
| Метод включает в себя двухэтапное обнаружение насилия и лиц в видео с использованием дескриптора VIF (variance inflation factor - коэффициент инфляции дисперсии) и алгоритмов нормализации [58] | VIF метод распознавания возражений CUDA и комплект детектора лица | Метод Horn-Schunck для гистограммы | Интерполяционная классификация | Менее многолюдно | Более низкая частота кадров 14% слишком высокая скорость 35 ФС/с 97% |

Продолжение таблицы 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 |
| SVM-метод распознавания, основанный на статистической теории без декодирования видеокадров [59] | Метод векторной нормализации | Метод макроблоков для извлечения объектов | Движение области и дескриптор для классификации видео | Многолюдно | 96.1% |
| Обнаружение боев с движущимися кадрами [60] | Бинаризация изображений | Пространственно-временной метод извлечения сгустков | Классификация на основе длины блоба, наибольший рассмотрите боевой | Многолюдно | 70% - 98% |
| Лагранжевы поля направления и напрашивается словесная основа для признания насилия в видео [61] | Глобальная компенсация движения объекта | Лагранжева теория и ступенчатый метод извлечения признаков движения | Поздний сплав для классификации | Многолюдно | 91% - 94% |
| Простой подход к формированию видео для предварительной обработки с последующим извлечением признаков и распознаванием нормального и аномального события [62] | Гауссова Модель | Примените различные формулы к последовательному кадру, чтобы извлечь необходимый объект | Классификация на основе правил с использованием порога по умолчанию | Менее многолюдно | 90% |
| Совершенно другая техника, в которой учится для борьбы с животными обнаруживать человеческую борьбу [63] | Область движения и метод оптического потока | Методы Vif, Vif | Трансфертный подход к обучению | Менее многолюдно | 90% |
| Фреймворк для обнаружения насилия устанавливает параметр, который насильственные люди просто анализируют в социальных сетях [64] | Подход к демографическому анализу | Этническая структура | Модели машинного обучения | Менее многолюдно | 95% |

1) Быстрое обнаружение драки

В работе [56] для выявления последовательностей насилия предложен следующий метод. Предполагается, что в сценах драки движущиеся кадры имеют определенную форму и положение. Во-первых, для абсолютных изображений вычисляется разница между последовательными кадрами. Затем полученное изображение бинаризуется, приводя к числу движущихся скоплений и помечая самый большой из них в последовательности драки и на сцене без драки. Выбираются только K самых больших сгустков движения. Чтобы классифицировать k больших двоичных объектов, вычисляются различные параметры, такие как центроид, площадь, периметр и расстояние между большими двоичными объектами. Затем капли характеризуются как борьба и не-борьба. Эксперименты, проведены с использованием данных фильмов (около 200 клипов), набора данных с хоккейными матчами, в которым есть 1000 роликов, и набора данных УКУ-111, состоящего из настоящих роликов, собранных с ютуба. Он имеет значительно более быстрое вычислительное время, что делает его приемлемым для приложений в режиме реального времени.

2) Вращательно-инвариантный признак моделирования когерентности движения (RIMOC)

События агрессии трудно поддаются определению из-за отсутствия последовательности и часто нуждаются в интерпретации высокого уровня. Поэтому принято классифицировать то, что часто присутствует в видеороликах с насильственным поведением людей на низком уровне, то есть неструктурированные и отрывистые движения. Для достижения этой цели предлагается инновационная задача-специфическое вращательно-инвариантное моделирование когерентности движения (RIMOC) [57]. Этот метод был создан на собственных значениях, полученных из гистограммы векторов оптического потока (HOF) из моментов последовательного временного, плотно и локально вычисленного и, кроме того, встроенного в сферическое Риманово многообразие. Этот метод используется для изучения моделей статистики в еженедельном контролируемом режиме. Многомасштабная схема, применяемая на основе метода вывода, позволяющего воспринимать события с неустойчивым движением во времени и пространстве, как хорошие кандидаты на насильственные события. Нет набора данных, доступного специально для агрессивных событий.

3) Быстрое распознавание лиц

Для достижения цели обнаружения лиц в сценах насилия в помощь средствам контроля безопасности предлагается метод быстрого обнаружения лиц [58]. Дескриптор насильственного потока (ViF – variance inflation factor) используется вместе с методом Horn-Schunck для обнаружения сцен насилия на ранней стадии. Затем для улучшения качества видео применяется алгоритм неадаптивной интерполяции суперразрешения. Наконец, срабатывает детектор лица Kanade-Lucas-Tomasi (KLT). Чтобы достичь очень быстрой обработки, детектор лиц и алгоритм суперразрешения параллельны CUDA. CUDA состоит из функций, которые выполняются одновременно в многочисленных легких потоках на графическом процессоре. Эксперименты проводятся для оценки предложенного метода с использованием набора данных Boss, а другой набор данных строится под названием violent dataset. Многообещающие результаты достигаются при обнаружении граней с точки зрения площади под кривой (AUC – area under the curve) и точности.

4) Распознавание насильственной активности без декодирования

Для целенаправленного обнаружения движений и их отслеживания в большинстве методов распознавания активности характерны сложность и ограниченность применения таких методов. Именно поэтому предлагается быстрый метод распознавания насильственной активности, основанный на векторах движения [59]. Во-первых, векторы движения извлекаются непосредственно из сжатых видеопоследовательностей. Во-вторых, атрибуты векторов движения анализируются в каждом кадре и между кадрами и достигают дескриптора вектора движения области (RMV). Наконец, радиальная основа была взята с использованием SVM в качестве функции ядра для классификации RMV и изучения того, присутствует ли активность насилия или нет. Существует множество наборов данных, доступных для распознавания активности, но ни один из них не подходит непосредственно, поскольку они сосредоточены на простом индивидуальном распознавании активности. Именно поэтому для оценки предложенного метода строится набор данных VVAR10, состоящий из 296 положительных и 277 отрицательных выборок. Образцы взяты из YouTube, UCF50, UCF sports и HMDB51. Эксперименты проводятся с использованием набора данных VVAR10, и результаты показывают, что он может обнаруживать 96,1% насильственных действий в видеопотоках, а скорость расчета быстра с точки зрения точности, частоты пропущенных сигналов тревоги (MAR) и частоты ложных сигналов тревоги (FAR), поэтому он подходит для встраиваемых систем.

5) Автоматическое обнаружение драки

Для обнаружения драк малозатратным и естественным способом предлагается подход, основанный на анализе движения [60]. Для обнаружения драки применяются два подхода. Первый - это двухуровневая статистическая агрегация, которая генерирует набор признаков. Из серии кадров путем вычисления векторов оптического потока извлекаются пиксели движения, а затем области движения. Затем изображения оптического потока классифицируются по характеру векторов после устранения помех. После этого статистика движения вычисляется в соответствии с классифицированными типами для получения набора признаков для распознавания. Второй используемый подход - это «сумка слов», которая используется для создания визуального набора слов. Затем гистограмма используется по всему набору визуальных слов в качестве вектора для характеристики видео для обнаружения драки. Эксперименты проводятся с использованием видеозаписей, основанных на реальных драках, и результаты эксперимента показывают, что предложенный метод превосходит существующие методы, основанные на дескрипторах MoSIFT с луковым механизмом и базовом анализе сигналов движения с луковым подходом с точки зрения отзыва и точности.

6) Обнаружение множественной аномальной активности

В работе [62] для решения задачи обнаружения подозрительных действий предлагается метод, который автоматически обнаруживает различные аномальные активности в видеоклипах. Эта система включает в себя три основных этапа: обнаружение движущегося объекта, отслеживание объекта и понимание поведения для распознавания активности. На первом этапе предварительной обработки обнаруживаются движущиеся объекты и производится удаление шума. Затем процесс извлечения объектов используется для определения ключевых объектов, таких как направление, скорость, размеры и центроид. Извлеченные функции помогают отслеживать объекты в видеокадрах. На последнем этапе метод классификации на основе правил используется для классификации действий из видео, и если обнаруживается какая-то подозрительная активность, она генерирует сигнал тревоги. Эксперименты проводятся на вновь созданном наборе данных, основанном на 45 видео, которые содержат три вида деятельности: бег, ходьба и ползание. Результаты экспериментов показывают, что предложенный метод позволяет обнаруживать различные виды аномальной активности в различных сценариях и дает наилучшие результаты с точки зрения точности.

7) Обнаружение насилия толпы

В приложениях компьютерного зрения теория Лагранжа предоставляет богатый набор инструментов для анализа долгосрочной нелокальной информации о движении. На основе этой теории предложена специализированная лагранжева методика [61] для автоматического распознавания сцен насилия в видеопоследовательностях. Пространственно-временная модель на основе Лагранжевых полей направления используется для новых объектов и использует информацию о компенсации фонового движения, появлениях и долговременном движении. Расширенный подход применяется в позднем слиянии на основе каждого видео в качестве классификационной схемы для обеспечения соответствующего временного и пространственного масштаба объектов. Эксперименты проводятся для оценки предложенного метода с использованием трех эталонных наборов данных, которые представляют собой хоккейную драку, насильственную толпу и насилие в фильмах. Результаты показывают, что добавление теории Лагранжа является ценным признаком для обнаружения насилия, а эффективность классификации повышается по сравнению с современными методами, такими как ViF, HOG + BoW, two stream CNN и т. д. С точки зрения AUC и точности.

8) Обнаружение межвидовой борьбы

В социальной обработке сигналов обнаружение поведения людей в виде драк на видео имеет важное значение, особенно в контексте наблюдения. В реальной жизни сбор данных для обнаружения драк обычно ограничивает возможности машинного обучения и влияет на производительность современных методов, основанных на данных. Для решения этой проблемы представлен инновационный метод межвидового обучения наряду с набором малозатратных вычислительных характеристик движения для обнаружения драки [65]. Это позволяет эффективно избежать проблемы ограниченных данных о человеческой борьбе. Предложенный метод использует существенную общность между драками животных и человека, такую как физическое ускорение движущихся частей тела. Предлагаемый метод берет входные данные из видеозаписей борьбы животных и нескольких видеороликов борьбы людей. Предложен набор локальных признаков движения (LMF), основанный на статистике движения, сегментной корреляции, следующей парадигме анализа движения. LMF извлекаются из каждого видео. Временные характеристики, основанные на человеческой эвристике, извлекаются и для обнаружения драки принимаются традиционные алгоритмы машинного обучения, такие как SVM. Ансамблевые классификаторы предлагаются для выполнения межвидового обнаружения борьбы. Эксперименты проводились с использованием образцов видеоклипов, наборов данных хоккея и фильмов. Результаты экспериментов показывают, что предложенный метод работает лучше, чем современные методы, такие как ViF, OViF и сигналы движения с точки зрения точности.

2.1.3 Методы обнаружения насилия с использованием SVM

В данном пункте отчета подробно обсуждаются методы обнаружения насилия, использующие в качестве классификатора машину опорных векторов (SVM – Support vector machine).

В таблице 2 представлен перечень методов распознавания насильственных событий. SVM – это алгоритм, который используется для решения задач классификации с использованием обучения с подкреплением. Данные в SVM строятся на размерном пространстве (числовые объекты) и дифференцируются в пределах двух классов. SVM - широко используемый метод в компьютерном зрении, так как он надежен и учитывает числовые характеристики. Он используется для задач, связанных с бинарной классификацией. SVM основан на ядре. Ядро - это функция, которая преобразует входные данные в высокомерное пространство, где решается задача. Основным недостатком SVM является недостаточная прозрачность результатов [66].

Таблица 2. Методы обнаружения насилия с использованием SVM

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Метод | Метод обнаружения объекта | Метод извлечения признаков | Тип происшествия | Точность % |
| 1 | 2 | 3 | 4 | 5 |
| Обнаружение насилия в переполненных сценах в режиме реального времени [67] | Дескриптор ViF | Набор функций | Многолюдный | 88% |
| Фреймворк Bag of words, использующий ускорение для обнаружения действий [68] | Алгоритмы фонового вычитания | Метод оценки эллипса для последовательных кадров | Менее многолюдно | Приблизительно 90% |
| Генетический алгоритм фреймворка с модулем отслеживания и обнаружения [69] | Гауссова модель | Алгоритм для извлечения оптического потока | Многолюдно | 82%-89% |
| Мультимодельные особенности фреймворка на базе подкласса [70] | Изображение CNN и ImageNet | Google Net для извлечения функций | Менее многолюдно | 98% |
| Определение частоты действий насильственного назначения [71] | Пространственные пирамиды и сетки для обнаружения объектов | Методы пространственно временной сетки для извлечения объектов | Многолюдно | 96%-99% с использованием различных наборов данных |
| Обнаружение насилия с использованием ориентированного насильственного потока [72] | Оптический метод потока | Сочетание ViF и OViF дескриптор | Многолюдно | 90% |
| AEI и HOG объединили фреймворк для распознавания аномального события в визуальных движениях [73] | Метод AEI для вычитания фона | HOG и пространственно - временные методы извлечения признаков | Как многолюдно, так и менее многолюдно | 94%-95% |
| Фреймворк включает в себя предварительную обработку, обнаружение активности и извлечение изображений. Эта работа идентифицирует аномальное событие и изображение из данных [74] | Оптический поток и временная разность для обнаружения объекта CBIR метод извлечения изображений | Функция Гаусса для анализа видеофайлов | Менее многолюдно | 97% |
| Поздний метод слияния для временных слоев восприятия для обнаружения активности высокого уровня. Использование несколько камер от 1 до N. [75] | Метод вектора движения для идентификации с нескольких камер в двух измерениях | SGT MtPL метод | Менее многолюдно | 98% |

Продолжение таблицы 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| Двухканальная сверточная нейронная сеть для обнаружения в реальном времени [76] | ImageNet для обнаружения объекта | VGG-f модель для извлечения признаков | Многолюдно | 91%-94% |
| Решите проблему обнаружения, разделив цель по глубине и четкому формату с помощью Connect [77] | Обнаружение движения и модель Trof | BoW подход | Менее многолюдно | 96% |
| Метод мешка слов с использованием пространственно-временного метода для обнаружения аномалий в видео [78] | Представление сегментов и подсегментов | HOF и HOG для получения видеокадров | Менее многолюдно | 84%-91% |

1) Обнаружение нарушения насилия в переполненных сценах в режиме реального времени

Хотя использование камер видеонаблюдения популярно, но их эффективность вызывает сомнения. Для решения сложной задачи мониторинга насилия в местах массового скопления людей предлагается инновационный подход к обнаружению насилия в режиме реального времени [67]. Предлагаемый метод учитывает статистику изменения величины векторов потока во времени, рассматриваются две взаимосвязанные, но различные задачи: классификация насилия и выявление насилия. Основная цель состоит в том, чтобы обнаружить изменение от насильственного к ненасильственному поведению с наименьшей задержкой с момента, когда это изменение произошло. Бурные потоки (Vif) дескриптор используется для представления статистических данных, собираемых для нескольких последовательных кадров. Затем дескрипторы Vif классифицируются как насильственное или ненасильственное поведение с использованием линейного SVM. Эксперименты проводятся с использованием набора данных, основанного на видео наблюдения, собранных с YouTube посредством сбора данных с матчей по хоккею. Результаты экспериментов показывают, что предложенный метод работает лучше, чем существующие методы, такие как HOG, HOF и т.д., В зависимости от величины полей оптического потока только с точки зрения площади под кривой (AUC) и точности.

2) Быстрое обнаружение насилия

В компьютерном зрении проблема распознавания действий становится активной темой. Хорошо известный модуль bag-of-words(багаж слов) был использован в недавней работе, которая используется для распознавания борьбы. В этой схеме пространственно-временные признаки были взяты из видеокадров и использованы для классификации, а показатели точности 90% достигнуты для этой задачи. Вдохновленный результатами, которые говорят о том, что кинематические характеристики являются единственными дискриминантными для конкретных действий, инновационный метод, который использовал метод экстремального ускорения в качестве основного признака, предложил [68]. Чтобы эффективно оценить экстремальные характеристики, случайное преобразование применяется к последовательным кадрам видео. И использовал SVM и Adaboost в качестве классификатора. Были проведены эксперименты с использованием двух специальных наборов данных записей с хоккейных матчей и фильмов, и результаты показывают, что 12%-ное увеличение точности и работает лучше, чем современные универсальные методы распознавания действий, которые имеют такие функции, как масштабно-инвариантное преобразование признаков (SIFT) и масштабно-инвариантное преобразование признаков движения (MoSIFT) [79], а также предлагаемый метод в 15 раз быстрее.

3) Гауссова модель оптического потока (GMOF)

Для систем наблюдения обнаружение насилия становится актуальной темой, но изучается не так глубоко, как распознавание действий. Более ранние методы в основном фокусировались на обнаружении насилия и предпринимали мало усилий для определения места насилия. Вот почему была предложена надежная и быстрая система обнаружения и локализации насилия в местах наблюдения [69]. В рамках выделения областей-кандидатов насилия представлена Гауссовская модель оптического потока (GMOF). Модель адаптивно спроектирована как отклонение от нормального поведения толпы, обнаруженное в сценах. Затем для каждого видеообъема, который создается путем плотной выборки областей потенциального насилия, выполняется обнаружение насилия. Затем, чтобы отличить сцены насилия от ненасильственных сцен, был также предложен инновационный дескриптор под названием ориентационная гистограмма оптического потока (OHOF). Сначала обучающий модуль выбирает обучающие данные и извлекает дескриптор OHOF, а затем, используя линейный SVM, получает модель объекта. В конце, дескриптор сравнивается с обученной моделью SVM известных сцен насилия. Эксперименты проводятся с использованием трех сложных наборов данных. Результаты эксперимента показывают, что предлагаемый дескриптор работает лучше, чем предыдущие дескрипторы, такие как гистограммы ориентированных градиентов, гистограмма распределения оптического потока (Хоф), сочетание Hog и Хоф (ГНФ), MoSIFT и просеять в плане площади под кривой (AUC) и точность.

4) Обнаружение насилия в видео с помощью подклассов

Чтобы решить сложную проблему обнаружения насилия в видео, Ли и др. [70] специально фокусируется на сочетании мультимодальных функций путем добавления и использования подклассов, визуально связанных с насилием. Для достижения этой цели используется набор данных MediaEval 2015. Видео набора данных помечаются вручную по отношению к подклассам и приводят к 10 подклассам насилия, таким как кровь, пистолет, смерть и другие. Понятие подклассов взято из существующих исследований. SVM используется в качестве классификатора для обучения подклассов, и набор делится на два случайно непересекающихся набора, 70% для обучения и 30% для тестирования. Эта процедура также применяется к тестовому набору, и результаты показывают, что решение, основанное на подклассах, превосходит существующие методы, содержащие функции движения, такие как HOG, Motion Boundary Histogram (MBH) и HOF со средней точностью 0,303 и точностью 100 0,55 на средневековом наборе данных.

5) Распознавание и выявление человеческого насилия

Распознавание и выявление насилия становится важной темой для видеозаписей наружного наблюдения. Основная цель состоит в том, чтобы определить, происходит ли насилие. Во-первых, расширение улучшенных векторов Фишера (IFV) предлагается для видеоклипов [71]. Используются локальные объекты и их пространственно-временные положения, что позволяет репрезентировать видео. Затем для обнаружения насилия изучается популярный подход скользящего окна. Для ускорения подхода используется структура данных суммированной таблицы площадей и переформулируются формулы IFV. Во-вторых, локальные пространственно-временные характеристики извлекаются из видео с использованием улучшенных плотных траекторий (IDT). Затем видеопрезентация для каждого дескриптора вычисляется независимо, как HOG для представления видео с помощью IFV. Затем линейный SVM-классификатор используется для распознавания насилия, и в конце концов, используя подход быстрого скользящего окна, обнаруживается насилие. Обширная оценка проводится с использованием 4 современных наборов данных насильственных потоков, фильмов и хоккейных матчей. Набор данных Violence-Flow 21 используется для задачи обнаружения насилия. И результаты показывают, что предложенные подходы работают лучше по сравнению с существующими подходами, такими как ГНФ, Хоф.

6) Обнаружение насилия с помощью ориентированного насильственного потока

Спрос рынка на интеллектуальное обнаружение насилия постоянно растет с помощью камер видеонаблюдения, но все же это сложная область исследований. Во-первых, для практического обнаружения насилия в видео предлагается инновационный метод извлечения признаков, получивший название Oriented Violent Flows (OViF) [72]. В статистических ориентациях движения он в полной мере использует информацию об изменении величины движения. AdaBoost используется для выбора объектов, а затем SVM-классификатор обучается на выбранных объектах. Эксперименты проводятся на наборах данных базы данных Hockey и Violent-Flow для оценки полезности предложенного метода, и результаты показывают, что предложенный метод работает лучше, чем базовые методы: LTP и ViF с точки зрения точности и AUC. Во-вторых, принимаются стратегии комбинирования признаков и мультиклассификаторов, и достигаются выдающиеся результаты. Результаты эксперимента показывают, что использование комбинированных функций с AdaBoost и Linear-SVM позволяет повысить производительность по сравнению с современными методами на бенчмарке Violent-Flows.

7) Автоматическая система видеонаблюдения в режиме реального времени

В системе наблюдения обнаружение подозрительной деятельности играет жизненно важную роль. В среде ученых существует острая потребность в системе наблюдения, которая может надежно функционировать. Именно поэтому предлагается новая структура для системы видеонаблюдения в режиме реального времени, которая обнаруживает автоматически [74]. Работа разделена на три этапа. Фаза предварительной обработки включает в себя обнаружение аномальной активности человека и фазу извлечения изображений на основе контента. На этапе предварительной обработки все студенты должны зарегистрироваться перед началом курса обучения, а регистрация включает в себя сбор личных данных, и студент должен представить свою собственную фотографию для создания студенческого билета. Предложенная система нуждается в изображениях студентов также в различных условиях, таких как гнев, страх, печаль и т. д. получить точное описание в терминах поиска изображений на основе контента (CBIR). Эти записи сохраняются в базе данных CBIR для случая выявления аномальной активности студентов. Затем на следующем этапе изображение преобразуется в кадры. Алгоритм временной дифференцировки используется для обнаружения движущихся объектов, а затем с помощью функции Гаусса определяются области движения. Далее для распознанных объектов, которые являются человеческими или нечеловеческими, в качестве фильтра используется модель формы, основанная на уравнении омега. Человеческая деятельность классифицируется как нормальная и ненормальная деятельность с использованием SVM. В случае ненормальной деятельности человека, машина генерирует автоматическое предупреждение. Используя CBIR, он также вставляет метод извлечения распознанного объекта из базы данных для идентификации и распознавания объекта. Наконец, проведено программное моделирование с использованием MATLAB, и экспериментальные результаты показывают, что система достигает отслеживания, семантического обучения сцены и обнаружения аномалий в среде без участия человека.

8) Надежное распознавание аномальной активности человека

Для выявления любой ненормальной активности у пожилых людей и поддержки идеи качественной и независимой жизни предлагается надежная система распознавания ненормальной активности человека [73]. Вычисляя интеграцию векторов признаков, которые представляют собой гистограмму ориентированных градиентов (HOG) и моментов Цернике на средних энергетических изображениях (AEI), фреймворк структурируется для построения надежного вектора признаков. Компактное представление видеопоследовательностей обеспечивается формированием AEI без потери пространственно-временной информации. Взаимные размеры вектора признаков уменьшаются путем применения PCA, а для классификации видов деятельности применяется SVM. Предлагаемая работа оценивается на общедоступном 3D-наборе данных “Kinect Activity Recognition Dataset” (KARD) и “UR fall detection”. Эксперименты показывают, что результаты достигли 94% и 95,22% для набора данных UR fall и набора данных KARD с точки зрения средней точности распознавания (ARA) соответственно.

9) Новая структура для анализа деятельности высокого уровня

Предложена новая структура, основанная на позднем слиянии для анализа высокоуровневой активности с использованием мультинезависимых временных слоев восприятия [75]. Строятся два типа слоев восприятия, основанных на SVM и деревьях ситуационных графов (SGT). Структура состоит из трех этапов: многовременного анализа, многовременных слоев восприятия и позднего слияния. Результаты, полученные из слоев многовременного восприятия, сплавляются в счет активности на этапе позднего слияния. Для оценки этого подхода структура применяется к выявлению насильственных событий при визуальном наблюдении. Для экспериментов используются три известных набора данных: NUS-HGA, Behavior и некоторые видео с YouTube. Экспериментальные результаты показывают, что многовременная структура превосходит существующие одновременные структуры с точки зрения точности, показывая, что использование многовременного метода имеет преимущество перед одновременными методами.

10) Обнаружение насилия в реальном времени

Для надежного обнаружения действий насилия обычно недостаточно выбранных вручную функций. Поэтому модели на основе BI-каналы CNN и SVM для обнаружения насилия предлагается [76]. Предлагаемая модель состоит из трех частей: извлечение признаков, обучение SVM и слияние меток. Во-первых, структура Би-каналов CNN используется для извлечения двух особенностей. Первый признак-это исходный видеокадр, который используется для извлечения признаков внешнего вида, а второй вход-это разность соседних кадров, которая используется для разделения признаков движения. Затем внешний вид и движение принимаются в качестве классификатора для линейного SVM. В конце концов, результат обнаружения насилия достигается с помощью метода слияния меток, который объединяет информацию о движении и информацию о внешнем виде. Эксперименты проводятся для оценки предложенного метода с использованием наборов данных хоккейной борьбы и насильственной толпы. Результаты экспериментов показывают, что предложенный метод работает лучше, чем существующие методы типа HOG, HOF, MoSIFT, SIFT, Two-stream во многих реалистичных сценах с точки зрения точности.

11) Breaking down violence

Использование автоматических средств обнаружения насилия в видеозаписях имеет существенное значение для анализа камер видеонаблюдения и правоохранительных органов для поддержания общественной безопасности. Кроме того, это помогает защитить детей от попадания в неподходящий контент и помогает родителям принять лучшее решение о том, что их дети должны смотреть. Хотя это сложная проблема, так как определение насилия очень субъективно и широко. Вот почему обнаружение насилия в видеоклипах без присмотра человека является не только концептуальной, но и технической проблемой. Чтобы преодолеть эту проблему, предлагается идея насилия, используемая для CNN, разбивая ее на более конкретные и объективные части [77]. Во-первых, для изучения особенностей, непосредственно связанных с насилием, таких как кровь, взрывы, драки и т. д., используются независимые сети. Затем, чтобы описать насилие, используя такие признаки, отдельные классификаторы SVM обучаются для каждого понятия, а затем объединяют их позже в метаклассификацию. Также исследуется, как представить основанные на времени события в качестве сетевых входных данных для неподвижных изображений, поскольку многие изображения определяются в форме движения. Эксперименты проведены для оценки предложенного метода с использованием набора данных EvalMedia 2013 VSD. Результаты показывают, что концепция разбиения насилия на более мелкие концепции оказалась эффективным решением с точки зрения MAP@100 и AUC.

12) Автоматическое определение стилей ведения драки

Для классификации видеозаписей боевых искусств предложен метод распознавания [78]. На этапе предварительной обработки, чтобы очистить ненужные части видео, сначала обрезайте и сегментируйте соответствующий набор данных. Затем для извлечения объектов используются пространственно-временные точки интереса для обнаружения областей в видео, связанных с движением в последовательности кадров. Для классификации обучающий словарь всех дескрипторов строится и затем преобразуется в кластеры. Затем каждое видео представляется в виде гистограммы этих кластеров. Кроме того, KNN или SVM применяются в качестве классификатора к видео. Эксперименты проводились с использованием видеозаписей Олимпийских игр, включая Олимпийские игры 2012 года в Лондоне и юношеские Олимпийские игры 2014 года в Нанкине. То есть выбираются два основных класса видео: дзюдо и тхэквондо. Результаты экспериментов показывают, что предложенный метод с использованием SVM в качестве классификатора превосходит по точности существующие методы, использующие KNN и SVM с ядром RBF в качестве классификатора.

1.2.4. Методы обнаружения насилия с использованием глубокого обучения

Здесь подробно рассматриваются методы обнаружения насилия, использующие алгоритмы глубокого обучения. В таблице 3 представлен список методов распознавания, использующих классификацию на основе сверточных нейронных сетей (CNN, Convnet) [80]. Глубокое обучение основано на нейронных сетях. Этот метод используется для классификации насильственного распознавания на основе набора данных и извлеченных объектов с использованием сверточных слоев. Теперь методы обнаружения насилия, использующие алгоритмы глубокого обучения, подробно разрабатываются отдельно.

Таблица 3. Методы выявления насилия с использованием глубокого обучения

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Метод | Метод обнаружения объекта | Метод извлечения признаков | Тип происшествия | Точность % |
| 1 | 2 | 3 | 4 | 5 |
| Обнаружение насилия с помощью 3D CNN [81] | 3D свертка используется для получения пространственной информации | Метод обратного распространения | Многолюдно | 91% |
| Глубокая архитектура для распознавания мест [82] | VGG VLAD метод поиска изображений | Метод обратного распространения для извлечения признаков | Многолюдно | 87%-96% |
| Жестокие сцены, используя СNN и глубоко-аудио особенности [83] | MFB | CNN модель | Многолюдно | Приблизительно 90% |
| Обнаружение жестоких видео с помощью convLSTM [84] | CNN вместе с ConvLSTM | CNN модель | Многолюдно | Приблизительно 97% |

Продолжение таблицы 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 |
| Выявление агрессивного поведения человека и с учетом траектории и глубоко СNN [85] | Глубокий CNN | Метод оптического потока | Многолюдно | 98% |
| Методология реорганизации леса Хафа [86] | Обнаружение объектов с использованием пространственно - временных признаков | Метод MoSIFT для извлечения видеофайлов | Менее многолюдно | 84%-96% |
| Обнаружение насилия с использованием пространственно-временных функций с помощью 3D CNN [87] | Нейронная сеть передает передает CNN полученную информацию | 3D CNN | Многолюдно | Приблизительно 97% |

1) Обнаружение насилия с помощью 3D CNN

Типичные методы обнаружения драк полагаются на знания предметной области для построения сложных объектов ручной работы из входных данных. С другой стороны, глубокие модели могут непосредственно действовать. Модель обучается с использованием контролируемого обучения, а градиентный спуск используется вместе с методом обратного распространения [81]. Эксперименты проводятся с использованием хоккейного набора данных, и результаты показывают, что предложенный метод работает лучше, не полагаясь на созданные вручную функции с точки зрения точности.

2) Глубокая архитектура для распознавания места

Для преодоления проблемы распознавания крупномасштабных визуальных мест, в которых основной задачей является быстрое и точное определение места по определенному запросу фотографий, была предложена архитектура CNN на основе слабо контролируемого распознавания мест [82]. Предлагаемый метод имеет три принципа. Во-первых, разрабатывается архитектура на основе CNN, которая обучается сквозным средствам для непосредственного выполнения задачи распознавания места. NetVLAD является ключевым элементом этой архитектуры, это новый всеобъемлющий слой VLAD, который поощряется локально агрегированными векторами дескрипторов, в основном используемыми для извлечения изображений. VLAD layer подходит для обучения с использованием обратного распространения и охотно подключается к любой архитектуре, основанной на CNN. Во-вторых, чтобы получить параметры архитектуры в сквозном средстве из изображений, представляющих одни и те же места в течение всего времени, которое берется из Google Street View Time Machine, разрабатывается недавно разработанная слабо контролируемая процедура обучения ранжированию на основе потерь. Наконец, эксперименты проводятся на основе типичного процесса оценки распознавания мест с использованием наборов данных Питтсбурга и Токио 24/7 в свободном доступе, и результаты показывают, что предлагаемая архитектура работает лучше, чем ненаучные представления изображений и готовые дескрипторы CNN на двух сложных тестах распознавания мест и современном представлении изображений на тесте поиска изображений.

3) Обнаружение жестоких сцен, используя СNN и глубоких нейронных сетей

Предложена система обнаружения насильственных сцен [83], использующая CNN, построенную на акустической информации из видеоклипов. Телеканал CNN подает информацию о звуке в двумя образами: в качестве классификатора и как экстрактор. Во-первых, 40-мерный Mel Filter-Bank (MFB) используется в качестве входного объекта для CNN с их Дельта-и. Затем видео преобразуется в короткие куски. Функции MFB разделены на 3 канала функций для изучения местных особенностей. После этого CNN используется для представления объектов. Функции на основе CNN используются для построения SVM-классификаторов. Затем обнаружение сцены насилия выполняется на каждом фрагменте видео. Далее обнаружение производится максимальным или минимальным объединением на уровне сегментов обнаружений. Эксперименты проводятся с помощью датасета MediaEval, и результаты показывают, что предлагаемый метод работает лучше, чем базовые методы: только аудио, только визуальный и аудио-выученный слияние и визуальный с точки зрения средней точности.

4) Обнаружение жестоких видео с помощью сверточной долговременной кратковременной памяти (ConvLSTM)

Для распознавания насилия в видео предложен метод на основе глубоких нейронных сетей [84]. CNN используется для извлечения функций из уровня кадра в видео. Затем эти функции накапливаются с помощью варианта LSTM, который использует сверточные ворота. Комбинация CNN и ConvLSTM может принимать локализованные пространственно-временные характеристики, что позволяет проводить локальный анализ движения, происходящего в видео. Также предлагается использовать разности соседних кадров в качестве входных данных для модели, которая кодирует изменения, произошедшие в видео. Эксперименты проводятся с использованием трех популярных наборов данных: хоккей, фильмы и видео насильстввенного характера. Результаты показывают, что предложенная модель работает лучше, чем современные методы, такие как ViF+Vif, ViF, three streams + LSTM и другие с точки зрения точности.

5) Обнаружение насильственного поведения человека путем интеграции траектории и глубокого CNN

Типичные методы обнаружения насилия обычно зависят от особенностей ручной работы, которая в основном не подходит. Вдохновленный работой глубоких моделей для распознавания человеческих действий, предложен инновационный метод обнаружения насильственного поведения человека путем объединения траектории и глубокого CNN [85], который использует преимущества как созданных вручную функций, так и глубоко изученных функций. Эксперименты проводятся на двух реальных наборах данных: хоккейная драка и насилие толпы. Полученные результаты показывают, что предложенный метод работает лучше, чем существующие методы: HOG, HOF, ViF и другие с точки зрения точности.

6) Метод распознавания драки

В компьютерном зрении распознавание действий становится важным направлением исследований. Такие задачи, как агрессивное поведение или драки, сравнительно мало изучены, но может быть полезно во многих сценариях видеонаблюдения, таких как тюрьмы, психиатрические палаты или личные смартфоны. Их обширная практичность создает интерес к разработке детекторов насилия или борьбы. Основным аспектом детекторов является эффективность, то есть эти подходы должны быть быстрыми в вычислительном отношении. Созданные вручную пространственно-временные характеристики достигают высокой точности как для внешнего вида, так и для движения, но извлечение некоторых характеристик все еще невозможно для реального применения. Впервые парадигма глубокого обучения применяется к задаче, использующей 3D CNN, которая принимает полную видеопоследовательность в качестве входных данных. Но особенности движения человека для этой задачи имеют решающее значение, и использование полного видео в качестве входного сигнала вызывает шум и избыточность в процессе обучения. Для этой цели был предложен гибридный фреймворк “handcrafted/learned” [86]. Метод, во-первых, направлен на получение иллюстративного изображения из видеопоследовательности, взятой в качестве входных данных для извлечения признаков, а лес Хафа используется в качестве классификатора. Затем, чтобы классифицировать это изображение и получить заключение для последовательности, используется 2D CNN. Проведены эксперименты с использованием трех хорошо известных наборов данных записей игр по хоккею, фильмов и сцен насилия. И результаты показывают, что предложенный метод работает лучше, чем различные методы ручного и глубокого обучения, основанные на точности и стандартных отклонениях.

7) Обнаружение насилия с использованием пространственно-временных функций с помощью 3D CNN

Для признания насильственных действий в целях безопасности обязательна усиленная система наблюдения, позволяющая избежать социального, экономического и экологического ущерба. Для этой цели предлагается структура трехэтапного сквозного обнаружения насилия в глубоком обучении [87]. Во-первых, в потоках видеонаблюдения люди обнаруживаются с помощью облегченной модели CNN, чтобы преодолеть и уменьшить огромную обработку непригодных кадров. Во-вторых, порядка 16 кадров с обнаруженными особями передаются в 3D CNN, где пространственно-временные характеристики этих последовательностей извлекаются и подаются в Классификатор Softmax. Затем 3D-модель CNN оптимизируется с помощью инструментария оптимизации нейронных сетей и открытого визуального вывода, разработанного Intel. Обученная модель преобразуется в промежуточную иллюстрацию и изменяет ее для исполнения на конечной платформе для окончательного обнаружения насилия. После обнаружения насилия сигнал тревоги передается в соседний отдел безопасности или полицейский участок, чтобы произвести действие. Эксперименты проводятся с использованием датасета драк толпы, хоккейных матчей и насилия в фильмах. Результаты эксперимента показывают, что предложенный метод работает лучше, чем современные методы, такие как ViF, AdaBoost, SVM, Hough Forest и 2D CNN, sHOT и другие с точки зрения точности и AUC.

2.2 Обзор существующих программных решений по распознаванию и классификации действий на видео

Действие тела также является критическим интерфейсом, привлекая все больше и больше внимания в последние годы. В этом разделе мы обобщаем исследования, связанные с действием тела, и обсуждаем следующие направления: распознавание действий, прогнозирование действий, непрерывный анализ действий, безопасность и анализ эгоцентрического видения действий.

Action Recognition: Визуальное распознавание действий является основным методом в распознавании буллинга. С помощью записи траекторий [88] отслеживали монокулярные и изученные траектории движения человека для распознавания действий. С развитием глубокого обучения были разработаны различные глубокие сети для получения высокой производительности распознавания для приложений HRI. [89] понимали намерения человека с помощью двух контролируемых рекуррентных нейронных сетей множественных временных масштабов (MTRNN). Первый слой распознает человеческие действия, а второй слой предсказывает человеческие намерения, основываясь на результатах первого слоя. Кроме того, [90] изучили особенности действия иерархическим образом, чтобы понять спонтанные эмоции для естественного эмоционального взаимодействия между человеком и роботами. Этот подход применил многоканальную сверточную нейронную сеть (CCNN) для интеграции мультимодальных функций распознавания невербальных эмоций, включая мимику и движения тела.

Информация о глубине полезна для описания действий человека в трехмерных пространствах. Как и традиционные подходы на основе RGB-видео, траектория также эффективна для описания действий с помощью 3D-информации о действиях. [91] пересекли суставы тела в направлениях и записали траектории этих направлений в последовательности действий. Они моделировали траектории и представляли активность в виде гистограммы векторов направления для интерактивного и неинтерактивного распознавания действий. Поскольку легко обнаружить суставы и части тела с захваченными скелетами, были предложены подходы к изучению и присвоению Весов важности суставам тела для эффективного представления действий [92]. Piyathilaka et al. [93] включили веса важности для 3D-скелетных суставов, чтобы игнорировать запутанные или нерелевантные функции, и обучили динамические байесовские сети (DBN) для классификации человеческой деятельности, чтобы еще больше улучшить возможности домашних сервисных роботов. Для представления действия были разработаны различные эффективные совместные функции. [94] представили Контрастивную модель распределения признаков (CFDM) в человеческих взаимодействиях, которая ввела внутрикадровый скелет и изучила дискриминационное представление для распознавания взаимодействия человека и человека. Кроме того, слияние признаков представлено для решения задачи распознавания сложных действий [95]. Charalampous et al. [96] позволили роботу ориентироваться в населенной среде, чтобы продемонстрировать социально приемлемое поведение. Он распознал человеческие действия с помощью RGB-видео и последовательностей глубины. Комбинируя информацию о глубине с информацией RGB, он способен повысить точность распознавания действий и способствовать применению действий в распознавании буллинга и насилия.

Взаимодействие человека и объекта-еще одно часто обсуждаемое направление. Поскольку она включает в себя человеческие действия и объекты, графовая модель, естественно, используется для представления взаимодействия. Коппула и др. [97] приняли структуру графа для моделирования взаимодействий человека и объекта, где узлы представляли объекты и под-действия, а ребра представляли отношения между объектами, отношения между объектами и под-действиями и их эволюцию во времени. Наконец, была представлена структурная векторная машина поддержки (SSVM) для распознавания взаимодействий. Кроме того, иерархическая обработка также представлена для решения этой проблемы. Granata et al. [98] представили двухслойный подход, состоящий из слоя восприятия и слоя классификации, для распознавания человеческой деятельности. На уровне восприятия траектории суставов скелета использовались для представления виртуальной динамики человека. Классификация слоя используется многоступенчатая модель опорных векторов (MSVM) признать объект участвует вспомогательных действий. Дискретный скрытый Марков был использован для распознавания взаимодействия. Кроме того, Dutta et al. [99] использовали вероятностную машину состояний (PSM) для моделирования человеческих поз, объектов и их отношений для распознавания действий. Эти подходы моделировали человеческие действия, объекты и их отношения для распознавания взаимодействия, имитируя реальные взаимодействия в HRI. Поэтому они обеспечили теоретическую подготовку к применению HRI.

3 Разработка требований для сбора данных

3.1 Разработка требований к процессу сбора данных

Прототип программно-технического комплекса будет апробироваться в школах образовательной системы РК. Видеоконтент, полученный с камер наблюдения, расположенных внутри здания (в коридорах, столовых и т.д.), будет последовательно проанализирован системой с целью выявления происшествий и определения их вида. При регистрации визуальных сигналов физической агрессии система распознавания лиц будет запоминать участников произошедшего инцидента, распределять их по ролям буллинга (“инициатор буллинга” и “жертва”) и фиксировать повторные происшествия с их участием. Данный видеоматериал будет просматриваться психологом с целью усовершенствования работы алгоритма, делая его способным оценивать эмоции участников буллинга по позам, жестам и движениям. Далее все просмотренные данные будут классифицироваться и маркироваться для обучения алгоритмов и адаптироваться для распознавания буллинга в режиме реального времени.

3.2 Разработка требований к качеству данных и их объему

Планируется создать набор данных для обучения объемом минимум 1000 часов непосредственных видео. Это реальная цифра, так как школ в государстве достаточно, так же достаточно инцидентов (обсуждаемые новости драк, избиения и т.д.). Как результат, набор данных должен состоять из видео, где можно будет видео идентифицировать лицо, классифицировать возраст, нападающего(их) и жертвы(в), «говорить» какого характера буллинг происходит (физический, социальный, вербальный).

Если говорить о конкретных критериях, то измеримыми критериями будут:

Целостность данных. Целостность данных или общность демонстрирует насколько совершенными и безошибочными представляются материалы как в доле коннотационного (смыслового) противоречия, так и в доле соотношения установленной текстуре или выбранного формата. Целостность зависит от корректного дробления данных на его составляющие.

Привязка к источнику данных: Привязка к источнику данных в данном контексте обозначает, что набор данных должен пополняться новыми данными для улучшения алгоритмов ИИ и нахождения все более мелких деталей при тренировке. И данные, которые устарели, следует стирать по причине ненадобности для освобождения хранимого места.

Измеримость данных

Измеримость данных нужна для корректного выполнения последующих процедур, анализа и обучения. Измерение общего объема данных необходимо для выбора инструментов, которыми мы будем пользоваться для их дальнейшей обработки(препросцессинга) и использования в моделях.

3.3 Разработка требований к разметке собранных данных

Для обучения моделей искусственного интеллекта необходима разметка видео 3 видов:

* временные отметки начала и завершения изучаемых инцидентов потенциального буллинга и/или агрессивного поведения;
* пространственные метки (выделение прямоугольниками) людей в кадре во время изучаемых инцидентов;
* пространственные метки (выделение прямоугольниками) лиц людей в кадре во время изучаемых инцидентов.

Изучаемые видео материалы должны быть разбиты на части длиной до 1 часа с назначением каждой части уникального идентификатора. Каждая часть должна быть просмотрена как минимум 3 раза с целью осуществления одного из видов разметки. Осуществлять разметку разных видов в одно и то же время не рекомендуется по требованиям качества. Временная разметка должна производиться с точностью до одной секунды.

Расположение людей на видео по отношению к камере (лицом, спиной, боком).

ЗАКЛЮЧЕНИЕ

В ходе проделанной работы был сделан анализ существующих алгоритмов машинного обучения и ИИ для автоматического выявления и классификации физического, социального и другого рода буллинга, в интернет-пространстве. Был проделан сравнительный анализ этих алгоритмов на примерах других работ и исследований зарубежных специалистов.

Кроме этого, был составлен список требований для создания собственного набора данных, специфичного характера. А именно записи со школьных камер наблюдений для тренировки алгоритмов на задачу выявления буллинга непосредственно в школьном помещении. Конкретно были выявлены 3 критерия для этой задачи: целостность данных, привязка к источнику данных, измеримость данных.

Целостность данных - это схема построения набора данных, которая обеспечивает условие того, что данные полны, условие того, что данные не были изменены при выполнении любой операции над ними, будь то передача, хранение или представление. Соответственно, для увеличения эффективности алгоритмов ИИ нам требовались целостные данные.

Привязка к источнику данных – критерий, который определяет, что данные должны своевременно пополняться. Если этого не делать, то невозможно улучшить алгоритм, так как на старых, одних и тех же данных нельзя сделать тот же алгоритм лучше.

Измеримость данных или же когерентность – мы должны знать, насколько большим является набор данных, а именно в мерах измерения (кБ, МБ, ТБ), для выбора алгоритма, подходящий для размеров разных видов наборов данных.

Выполненная работа соответствует календарному плану. Поставленные задачи выполнены полностью.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 Alsaker F. D., Nagele C. Bullying in kindergarten and prevention // An international perspective on understanding and addressing bullying. — 2008. — № 1. — P. 230—252.

2 Salmivalli C., Kärnä A. Poskiparta E. From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program // Handbook of bullying in schools: An international perspective. — 2010. — № 1 — P. 441—454.

3 Olweus, D. Bullying at school: What we know and what we can do. — Malden, MA: Blackwell Publishing, 1993. — 140 p.

4 Ttofi M.M., Farrington D.P., Lösel, F., Loeber, R. The predictive efficiency of school bullying versus later offending: A systematic/meta-analytic review of longitudinal studies // Criminal Behaviour and Mental Health. — 2011. — № 21. — P. 80—89.

5 Olweus, D., Limber, S.P., Breivik, K. Addressing specific forms of bullying: A large-scale evaluation of the Olweus bullying prevention program // International Journal of Bullying Prevention. — 2019. — № 1(1). — P. 70—84.

6 Limber S.P. Development, evaluation, and future directions of the Olweus bullying programme // Journal of School Violence. — 2011. — № 10(1). — P. 71—87.

7 Olweus D., Limber, S.P. Bullying in school: Evaluation and dissemination of the Olweus bullying prevention program // American Journal of Orthopsychiatry. — 2010. — № 80(1). — P. 124–134.

8 Alsaker F.D., Nägele C. Bullying in kindergarten and prevention // An international perspective on understanding and addressing bullying. — 2008. — Vol. 1. — P. 230—252.

9 Salmivalli C., Kärnä A., Poskiparta E. From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program // Handbook of bullying in schools: An international perspective. — 2010. — № 1. — P. 441—454.

10 Raskauskas J.L., Gregory J., Harvey S.T., Rifshana F., Evans, I.M. Bullying among primary school children in New Zealand: Relationships with prosocial behaviour and classroom climate // Educational Research. — 2010. — № 52(1). — P. 1—13.

11 Goodwin J., Bradley S.K., Donohoe P., Queen K., O’Shea M., Horgan A. Bullying in schools: an evaluation of the use of drama in bullying prevention // Journal of Creativity in Mental Health. — 2019. — № 14(3). — P. 329—342.

12 Alsaker F.D., Nagele C. Bullying in kindergarten and prevention // An international perspective on understanding and addressing bullying. — 2008. — P. 230—252.

13 Raskauskas J.L., Gregory J., Harvey S.T., Rifshana F., Evans I.M. Bullying among primary school children in New Zealand: Relationships with prosocial behaviour and classroom climate // Educational Research. — 2010. — № 52(1). — P. 1—13.

14 Biggs B.K., Vernberg E.M., Twemlow S.W., Fonagy P., Dill E.J. Teacher adherence and its relations to teacher attitudes and student outcomes in an elementary school-based violence prevention program // School Psychology Review. — 2008. — № 37(4). — P. 533—549.

15 Brendgen M., Poulin F. Continued bullying victimization from childhood to young adulthood: A longitudinal study of mediating and protective factors // Journal of abnormal child psychology. — 2018. — № 46(1). — P. 27—39.

16 Kidziński Ł., Yang B., Hicks J.L., Rajagopal A., Delp S.L., Schwartz M.H. Deep neural networks enable quantitative movement analysis using single-camera videos // Nature communications. — 2020. — № 11(1). — P. 1—10.

17 Gracia I.S., Suarez O.D., Garcia G.B., Kim T.-K. Fast fight detection // PLoS ONE. — 2015. — № 10(4). — P. e0120448

18 Zhou P., Ding Q., Luo H., Hou X. Violence detection in surveillance video using low-level features // PLoS one. — 2018. — № 13(10). — P. e0203668.

19 Carneiro S.A., da Silva G.P., Guimaraes S.J.F., Pedrini H. Fight Detection in Video Sequences Based on Multi-Stream Convolutional Neural Networks // 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI). — 2019. — P. 8—15.

20 Febin I.P., Jayasree K., Joy P.T. Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm // Pattern Analysis and Applications. — 2020. — № 23. — P. 611—623.

21 Datta A., Shah M., Lobo N.D.V. Person-on-person violence detection in video data. // Object recognition supported by user interaction for service robots. — 2002. — № l. — P. 433—438.

22 Nievas E.B., Suarez O.D., García G.B., Sukthankar R. Violence detection in video using computer vision techniques // International conference on Computer analysis of images and patterns. — 2011. — P. 332—339.

23 Zhou P., Ding Q., Luo H., Hou X. Violent interaction detection in video based on deep learning // Journal of physics: conference series. — 2017. — № 844(1). — P. 12—44.

24 Pawar K., Attar V. Deep learning approaches for video-based anomalous activity detection // World Wide Web. — 2019. — № 22(2). — P. 571—601.

25 Ullah F.U.M., Ullah A., Muhammad K., Haq I.U., Baik S.W. Violence detection using spatiotemporal features with 3D convolutional neural network // Sensors. — 2019. — № 19(11). — P. 1—15.

26 Olweus D. Aggression in the schools: Bullies and whipping boys. — Hemisphere Publishing Corporation, 1978. — 218 p.

27 Fung A.L.C. Adolescent reactive and proactive aggression, and bullying in Hong Kong: prevalence, psychosocial correlates, and prevention // Journal of Adolescent Health. — 2019. — Vol. 64, №. 6. — P. S65-S72.

28 Jenkins L.N., Demaray M.K., Dorio N.B., Eldridge M. The Law and Psychology of Bullying // Advances in Psychology and Law. — Springer, Cham, 2019. — P. 197—234. (DOI: 10.1007/978-3-030-11042-0\_7)

29 Camodeca M. Bullying and victimization at school. Doctoral Thesis. — Amsterdam: PI Research, 2003. — 116 p.

30 Brito C.C., Oliveira M.T. Bullying and self-esteem in adolescents from public schools // Jornal de Pediatria. — 2013. — Vol. 89, № 6. — P. 601—607. (DOI: 10.1016/j.jped.2013.04.001)

31 Oliveira W.A.D., Silva M.A.I., Silva J.L.D., Mello F.C.M.D., Prado R.R.D., Malta D.C. Associations between the practice of bullying and individual and contextual variables from the aggressors' perspective // Jornal de Pediatria. — 2016. — Vol. 92, № 1. — P. 32—39. (https://doi.org/10.1016/j.jped.2015.04.003)

32 Chiebuka P., Ndukuba A., Abasiubong F. Aggressive behaviour among in-school adolescents in a developing country: patterns and associated factors // International journal of adolescent medicine and health. — 2020. — Vol. 1 (ahead-of-print). (https://doi.org/10.1515/ijamh-2020-0027)

33 Qiao-Zhi G.U.O., Wen-Jun M.A., Shao-Ping N.I.E., Yan-Jun X.U., Hao-Feng X.U., Zhang Y.R. Relationships between weight status and bullying victimization among school-aged adolescents in Guangdong Province of China // Biomedical and Environmental Sciences. — 2010. — Vol. 23, № 2. — P. 108—112.

34 Fatma A.A., Hanaa M.M. Bullying Behavior among School-Age Children and Its Relationship with Body Weight and Self-Esteem // Journal of Nursing and Health Science. — 2019. — Vol. 8, № 6, Ser. VIII. — P. 53—61.

35 Reulbach U., Ladewig E.L., Nixon E., O'Moore M., Williams J., O'Dowd T. Weight, body image and bullying in 9-year-old children // Journal of paediatrics and child health. — 2013. — Vol. 49, № 4. — P. E288—E293.

36 Pontillo M., Tata M.C., Averna R., Demaria F., Gargiullo P., Guerrera S., Pucciarini M.L., Santonastaso O., Vicari S. Peer victimization and onset of social anxiety disorder in children and adolescents // Brain sciences. — 2019. — Vol. 9, № 6. — P. 132. (https://doi.org/10.3390/brainsci9060132)

37 Esbensen F.A., Carson D.C. Consequences of being bullied: Results from a longitudinal assessment of bullying victimization in a multisite sample of American students // Youth & Society. — 2009. — Vol. 41, № 2. — P. 209—233. (doi: 10.1177/0044118X09351067)

38 Du Plessis M.R., Smeekens S., Cillessen A.H., Whittle S., Güroǧlu B. Bullying the brain? Longitudinal links between childhood peer victimization, cortisol, and adolescent brain structure // Frontiers in psychology. — 2019. — Vol. 9. — P. 2706. (https://doi.org/10.3389/fpsyg.2018.02706)

39 Swartz J.R., Carranza A.F., Knodt A.R. Amygdala activity to angry and fearful faces relates to bullying and victimization in adolescents // Social cognitive and affective neuroscience. — 2019. — Vol. 14, № 10. — P. 1027—1035. (https://doi.org/10.1093/scan/nsz084)

40 Nolfe G., Cirillo M., Iavarone A., Negro A., Garofalo E., Cotena A., Lazazzara M., Zontini G., Cirillo S. Bullying at workplace and brain-imaging correlates // Journal of clinical medicine. — 2018. — Vol. 7, № 8. — P. 200.

41 Sarzosa M., Urzúa S. Bullying among adolescents: The role of cognitive and non-cognitive skills. — National Bureau of Economic Research, 2015. — № w21631.

42 Fung A.L.C. Adolescent reactive and proactive aggression, and bullying in Hong Kong: prevalence, psychosocial correlates, and prevention // Journal of Adolescent Health. — 2019. — Vol. 64, № 6. — P. S65—S72.

43 Husky M.M., Delbasty E., Bitfoi A., Carta M.G., Goelitz D., Koç C., Lesinskiene S., Mihova Z., Otten R., Kovess-Masfety V. Bullying involvement and self-reported mental health in elementary school children across Europe // Child abuse & neglect. — 2020. — Vol. 107. — P. 104601.

44 Wolke D., Lereya S.T., Fisher H.L., Lewis G., Zammit S. Bullying in elementary school and psychotic experiences at 18 years: a longitudinal, population-based cohort study // Psychological medicine. — 2014. — Vol. 44, № 10. — P. 2199. (doi:10.1017/S0033291713002912)

45 Boden J.M., van Stockum S., Horwood L.J., Fergusson D.M. Bullying victimization in adolescence and psychotic symptomatology in adulthood: evidence from a 35-year study // Psychological Medicine. — 2016. — Vol. 46, № 6. — P. 1311—1320. (DOI: 10.1017/S0033291715002962)

46 Lereya S.T., Copeland W.E., Costello E.J., Wolke D. Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries // The Lancet Psychiatry. — 2015. — Vol. 2, № 6. — P. 524—531. (DOI: 10.1016/S2215-0366(15)00165-0)

47 Buch-Frohlich A., Paradis A., Hébert M., Cyr M., Frappier J.-Y. Bullying and sexual harassment as predictors of suicidality in sexually abused adolescent girls // International Journal Of Victimology. — 2019. — № 35. — P. 63—73.

48 Yang T., Guo L., Hong F., Wang Z., Yu Y., Lu C. Association between bullying and suicidal behavior among Chinese adolescents: an analysis of gender differences // Psychology research and behavior management. — 2020. — Vol. 13. — P. 89.

49 Bauman S. Do we need more measures of bullying? // Journal of Adolescent Health. — 2016. — № 59(5). — P. 487—488.

50 Solberg M.E., Olweus D. Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire // Aggressive Behavior: Official Journal of the International Society for Research on Aggression. — 2003. — Vol. 29, № 3. — P. 239—268.

51 Shetgiri R. Bullying and victimization among children // Advances in pediatrics. — 2013. — № 1. — P. 33.

52 Lloyd K., Rosin P.L., Marshall D., Moore S.C. Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures // Machine Vision and Applications. — 2017. — № 28(3—4). — P. 361—371.

53 Bilinski P., Bremond F. Human violence recognition and detection in surveillance videos // 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). — 2016. — P. 30—36.

54 Yan M., Meng J., Zhou C., Tu Z., Tan Y.P., Yuan J. Detecting spatiotemporal irregularities in videos via a 3D convolutional autoencoder // Journal of Visual Communication and Image Representation. — 2020. — № 67. — P. 1—11.

55 Olmos R., Tabik S., Herrera F. Automatic handgun detection alarm in videos using deep learning // Neurocomputing. — 2018. — № 275. — P. 66—72.

56 Gracia I.S., Suarez O.D., Garcia G.B., Kim T.-K. Fast fight detection // PLoS ONE. — 2015. — № 10(4). — P. e0120448.

57 Ribeiro P.C., Audigier R., Pham Q.C. RIMOC, a feature to discriminate unstructured motions: Application to violence detection for videosurveillance // Comput. Vis. Image Understand. — 2016. — № 144. — P. 121—143.

58 Arceda V.E.M., Fabián K.M.F., Laura P.C.L., Tito J.J.R., Cáceres J.C.G. Fast face detection in violent video scenes // Electron. Notes Theor. Comput. Sci. — 2016. — № 329. — P. 5—26.

59 Xie J., Yan W., Mu C., Liu T., Li P., Yan S. Recognizing violent activity without decoding video streams // Optik. — 2016. — № 127(2). — P. 795—801.

60 Fu E.Y., Va Leong H., Ngai G., Chan S. Automatic fight detection in surveillance videos // Proc. 14th Int. Conf. Adv. Mobile Comput. Multi Media. — 2016. — P. 225—234.

61 Senst T., Eiselein V., Kuhn A., Sikora T. Crowd violence detection using global motion-compensated Lagrangian features and scale-sensitive video-level representation // IEEE Trans. Inf. Forensics Security. — 2017. — № 12(12). — P. 2945—2956.

62 Chaudhary S., Khan M.A., Bhatnagar C. Multiple anomalous activity detection in videos // Procedia Comput. Sci. — 2018. — № 125. — P. 336—345.

63 Fu E.Y., Huang M.X., Va Leong H., Ngai G. Cross-species learning: A low-cost approach to learning human fight from animal fight // Proc. 26th ACM Int. Conf. Multimedia. — 2018. — P. 320—327.

64 Coletto M., Lucchese C., Orlando S. Do violent people smile: Social media analysis of their profile pictures // Proc. Companion Proc. Web Conf. — 2018. — P. 1465—1468.

65 Mabrouk A.B., Zagrouba E. Abnormal behavior recognition for intelligent video surveillance systems: A review // Expert Syst. Appl. — 2018. — № 91. — P. 480—491.

66 Auria L., Moro R. Advantages and disadvantages of support vector machines // Credit Risk Assessment Revisited: Methodological Issues and Practical Implications. — 2007. — P. 49—68.

67 Hassner T., Itcher Y., Kliper-Gross O. Violent flows: Real-time detection of violent crowd behavior // Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops. — 2012. — P. 1—6.

68 Deniz O., Serrano I., Bueno G., Kim T.-K. Fast violence detection in video // Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP). — 2014. — № 2. — P. 478—485.

69 Zhang T., Yang Z., Jia W., Yang B., Yang J., He X. A new method for violence detection in surveillance scenes // Multimedia Tools Appl. — 2016. — № 75(12). — P. 7327—7349.

70 Li X., Huo Y., Jin Q., Xu J. Detecting violence in video using subclasses // Proc. 24th ACM Int. Conf. Multimedia. — 2016. — P. 586—590.

71 Bilinski P., Bremond F. Human violence recognition and detection in surveillance videos // Proc. 13th IEEE Int. Conf. Adv. Video Signal Based Surveill (AVSS). — 2016. – P. 30—36.

72 Gao Y., Liu H., Sun X., Wang C., Liu Y. Violence detection using oriented violent flows // Image Vis. Comput. — 2016. — № 48. — P. 37—41.

73 Dhiman C., Vishwakarma D.K. High dimensional abnormal human activity recognition using histogram oriented gradients and zernike moments // Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC). — 2017. — P. 1—4.

74 Al-Nawashi M., Al-Hazaimeh O.M., Saraee M. A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments // Neural Comput. Appl. — 2017. — № 28(1). — P. 565—572.

75 Song D., Kim C., Park S.-K. A multi-temporal framework for highlevel activity analysis: Violent event detection in visual surveillance // Inf. Sci. — 2018. — № 447. — P. 83—103.

76 Xia Q., Zhang P., Wang J., Tian M., Fei C. Real time violence detection based on deep spatio-temporal features // Proc. Chin. Conf. Biometric Recognit. — 2018. — P. 157—165.

77 Peixoto B.M., Avila S., Dias Z., Rocha A. Breaking down violence: A deep-learning strategy to model and classify violence in videos // Proc. 13th Int. Conf. Availability, Rel. Secur. — 2018. — № 50 — P. 1—7.

78 Mishra A.A., Srinivasa G. Automated detection of fighting styles using localized action features // Proc. 2nd Int. Conf. Inventive Syst. Control (ICISC). — 2018. — P. 1385—1389.

79 Chen M.-Y., Hauptmann A. MoSIFT: Recognizing human actions in surveillance videos // Figshare. Journal Contribution. — 2018. — P. 1—16.

80 Agrawal T., Kumar A., Saraswat S.K. Comparative analysis of convolutional codes based on ML decoding // Proc. 2nd Int. Conf. Commun. Control Intell. Syst. (CCIS). — 2016. — P. 41—45.

81 Ding C., Fan S., Zhu M., Feng W., Jia B. Violence detection in video by using 3D convolutional neural networks // Proc. Int. Symp. Visual Comput. — 2014. — P. 551—558.

82 Arandjelovic R., Gronat P., Torii A., Pajdla T., Sivic J. NetVLAD: CNN Architecture for weakly supervised place recognition // Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR). — 2016. — P. 5297—5307.

83 Mu G., Cao H., Jin Q. Violent scene detection using convolutional neural networks and deep audio features // Proc. Chin. Conf. Pattern Recognit. — 2016. — P. 451—463.

84 Sudhakaran S., Lanz O. Learning to detect violent videos using convolutional long short-term memory // Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS). — 2017. — P. 1—6.

85 Meng Z., Yuan J., Li Z. Trajectory-pooled deep convolutional networks for violence detection in videos // Proc. Int. Conf. Comput. Vis. Syst. — 2017. — P. 437—447.

86 Serrano I., Deniz O., J. Espinosa-Aranda L., Bueno G. Fight recognition in video using Hough forests and 2D convolutional neural network // IEEE Trans. Image Process. — 2018. — № 27(10). — P. 4787—4797.

87 Ullah F.U.M., Ullah A., Muhammad K., Haq I.U., Baik S.W. Violence detection using spatiotemporal features with 3D convolutional neural network // Sensors. — 2019. — № 11. — P. 2472.

88 Green R.D., Guan L. Quantifying and recognizing human movement patterns from monocular video images-part i: a new framework for modeling human motion // IEEE Transactions on Circuits and Systems for Video Technology. — 2004. — № 14(2). — P. 179—190.

89 Hwang J., Tani J. Seamless integration and coordination of cognitive skills in humanoid robots: a deep learning approach // IEEE Transactions on Cognitive and Developmental Systems. — 2017. — № 10(2). — P. 345—358.

90 Marinoiu E., Zanfir M., Olaru V., Sminchisescu C. 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — P. 2158—2167.

91 El-Ghaish H., Hussein M.E., Shoukry A., Onai R. Human action recognition based on integrating body pose, part shape, and motion // IEEE Access. — 2018. — № 6. — P. 49040—49055.

92 Song S., Lan C., Xing J., Zeng W., Liu J. An end-to-end spatio-temporal attention model for human action recognition from skeleton data // Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. — 2017. — P. 4263—4270.

93 Arici T., Celebi S., Aydin A.S., Temiz T.T. Robust gesture recognition using feature pre-processing and weighted dynamic time warping // Multimedia Tools and Applications. —2014. — № 72(3). — P. 3045—3062.

94 Ji Y., Cheng H., Zheng Y., Li H. Learning contrastive feature distribution model for interaction recognition // Journal of Visual Communication and Image Representation. — 2015. — № 33. — P. 340—349.

95 Huynh-The T., Le B.V., Lee S., Yoon Y. Interactive activity recognition using pose-based spatio–temporal relation features and four-level Pachinko Allocation Model // Information Sciences. — 2016. — № 369. — P. 317—333.

96 Charalampous K., Kostavelis I., Boukas E., Amanatiadis A., Nalpantidis L., Emmanouilidis C., Gasteratos A. Autonomous robot path planning techniques using cellular automata // Robots and lattice automata. — 2015. — P. 175—196.

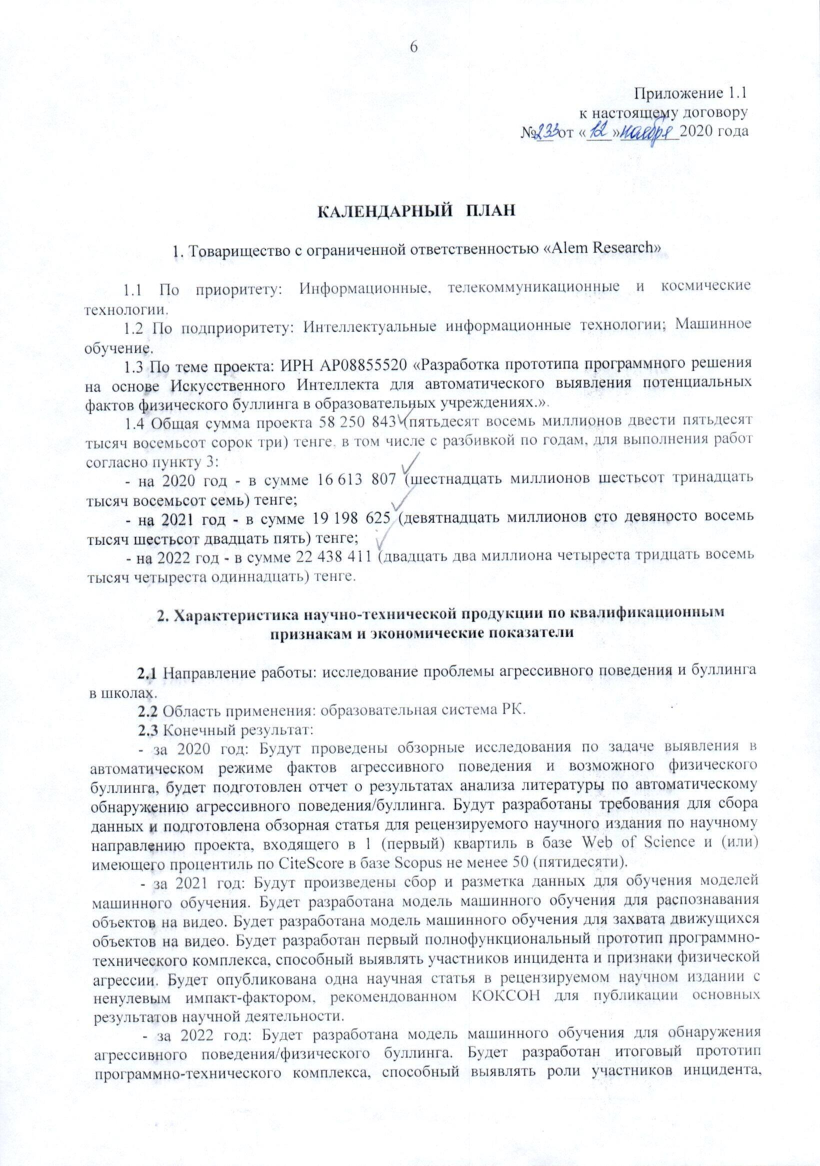
97 Koppula H.S., Gupta R., Saxena A. Learning human activities and object affordances from rgb-d videos // International Journal of Robotics Research. — 2013. — № 32(8). — P. 951—970.

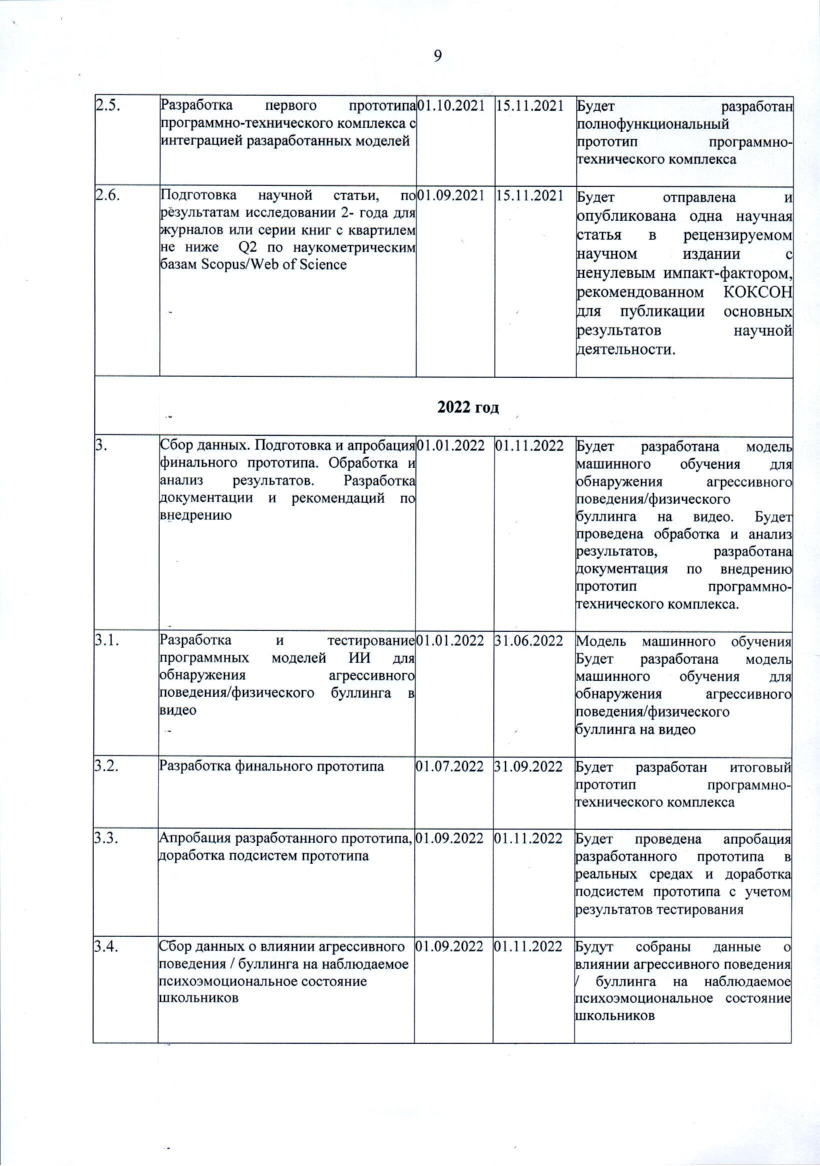
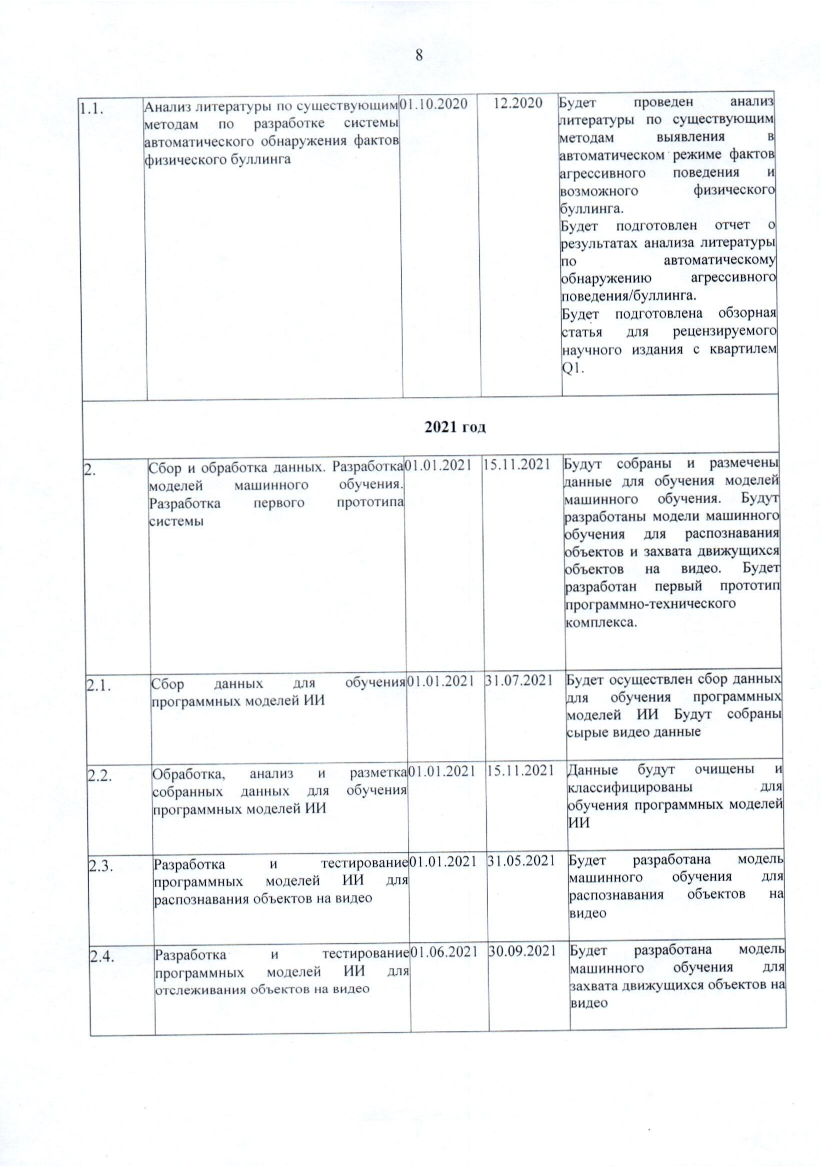
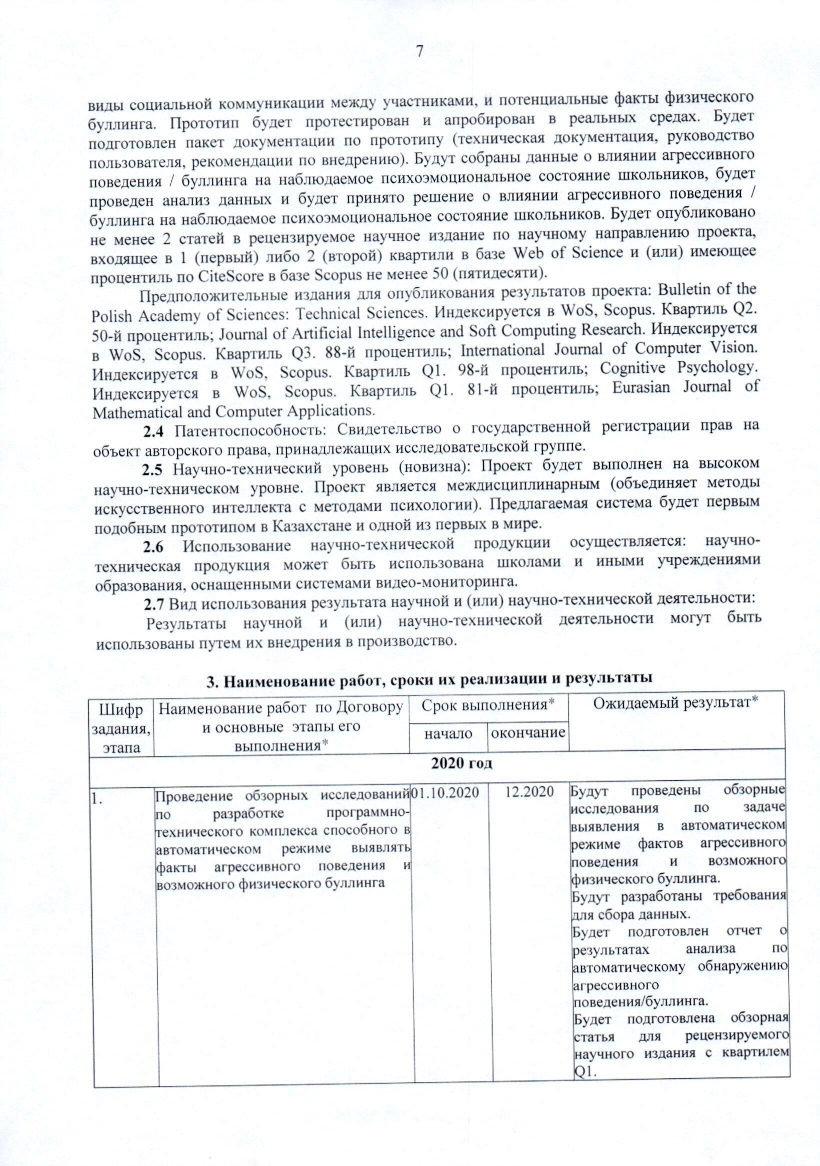
98 Granata C., Ibanez A., Bidaud P. Human activity-understanding: A multilayer approach combining body movements and contextual descriptors analysis // International Journal of Advanced Robotic Systems. — 2015. — № 12(7). — P. 889—897.

99 Dutta V., Zielinska T. Predicting the intention of human activities for real-time human-robot interaction (hri) // International Conference on Social Robotics. — Springer, Cham, 2016. — P. 723—734.

ПРИЛОЖЕНИЕ А

**Календарный план**





ПРИЛОЖЕНИЕ Б

**Обзорная статья**

**Обнаружение насилия в автоматизированном видеонаблюдении: Обзор**

**Аннотация:**

На данный момент существует растущий спрос на автоматизированные системы видеонаблюдения с широким спектром угроз в обществе и меньшим количеством рабочей силы для их мониторинга. Особенно трудно обнаружить насилие в переполненных местах из-за быстрого движения, перекрывающихся черт из-за загроможденного фона. В данной работе мы рассматриваем последние тенденции в выявлении насилия и проводим сравнительное исследование различных методик обнаружения такой деятельности по видеозаписи, предложенные в последние годы. В данной статье рассматриваются различные современные методы выявления насилия. В данной работе методы обнаружения делятся на три категории, основанные на классификации используемых методов: обнаружение насилия с использованием традиционного машинного обучения, использование машины опорных векторов (SVM) и использование глубокого обучения. Также представлены методы извлечения признаков и методы обнаружения объектов каждого отдельного метода. Кроме того, обсуждаются также наборы данных и видеоматериалы, используемые в техниках, которые играют жизненно важную роль в процессе распознавания. Для лучшего понимания последовательные этапы подходов были представлены на архитектурной диаграмме. Были обсуждены общие результаты исследований, которые могут быть полезны для поиска потенциальных будущих рабочих мест в этой области исследований.

**1. ВВЕДЕНИЕ**

Во всем мире растет число школьных программ по борьбе с хулиганством [1-2]; многие из них были разработаны на основе вмешательства, разработанного [3]. Однако эти программы имели разную степень успеха, от уменьшения до обострения проблемы [4]. Программы могут потерпеть неудачу из-за недостаточной точности развертывания инициативы, ограниченной поддержки учителей и персонала и/или отсутствия долгосрочного наблюдения.

Olweus Bullying Prevention Program (OBPS – Программа Ольвеуса для Профилактики Издевательств) была первой общешкольной программой по борьбе с хулиганством, разработанной в Норвегии [5]. Программа Ольвеуса включала в себя информацию о хулиганстве для учителей и родителей, записанное видео о школьных хулиганствах и анкету о жертвах хулиганства для учащихся. С момента завершения первоначального проекта OBPS разрабатывался и оценивался неоднократно [6]. Полученные результаты показали неизменно положительные эффекты [7].

К числу других примечательных программ относятся программа Be-Prox в Швейцарии [8], программа KiVa в Финляндии [9] и программа Kia Kaha в Новой Зеландии [10]. В Ирландии были введены такие программы, как Stay Safe [11]. Результаты оценочных исследований подчеркивают важное влияние этих программ на эмпатию [12-13]. Кроме того, принятие целостного школьного подхода рассматривается как важнейший компонент [14]. Брендген и Пулин [15] отметили, что программы профилактики издевательств выиграют от дополнительных вмешательств, специально направленных на развитие социальных навыков и создание поддерживающей сети.

Методы обнаружения насилия с помощью компьютерного зрения анализируют видео с камер наблюдения. В течение последних нескольких лет эти камеры и другое оборудование наблюдения устанавливаются в различных местах для обеспечения общественной безопасности, например в учебных заведениях, больницах, банках, на рынках, улицах и т.д., Чтобы следить за активностью людей [16]. Мониторинг включает в себя анализ поведения людей, независимо от того, являются ли их действия подозрительными или нормальными. Это обнаружение подозрительной активности в течение 24/7 или поиск такой активности в огромных данных, состоящих из записанных видео, является очень сложной задачей [17-20]. Для этого были разработаны различные методы распознавания человеческой деятельности в реальной жизни. Эти методы помогают обнаружить подозрительные действия на видеозаписях наблюдения.

Обнаружение насилия по видеозаписям наблюдения [21] также является одним из видов обнаружения активности. Разработано несколько методик и приемов для обнаружения событий жестокости и других вредных паттернов в видео [22-24]. В этих методах предлагаются различные подходы, которые работают с различными входными параметрами. Параметры-это в основном различные атрибуты или особенности видео, такие как ускорение, поток, время, внешний вид и т. д. В процессе обнаружения активности насилия первым шагом является разделение всего видео на сегменты и кадры [25]. Во-вторых, обнаруживает объект из видеокадров. В-третьих, извлечение особенности видео в соответствии с применяемым методом. Наконец, обнаруживает аномальную активность от кадров. Шаг варьируется в зависимости от метода, который применяется для обнаружения.

**2. КЛАССИФИКАЦИЯ ПРОЯВЛЕНИЙ БУЛЛИНГА В ШКОЛЬНОЙ СРЕДЕ**

Согласно определению Олвеуса [26], буллинг в школе - это поведение, при котором жертва буллинга многократно и в течение времени подвергается негативным действиям со стороны одного или нескольких школьников. Негативные действия могут быть в виде физического контакта, слов или других способов, например, изменения выражения лица или непристойные жесты, а также стремление преднамеренно исключать жертву буллинга из группы, сообщества. В целом буллинг характеризуется тремя критериями [27]:

* агрессивное поведение или преднамеренное «причинение вреда»;
* которое происходит «многократно и в течение времени»;
* имеется дисбаланс власти.

Буллинг подразумевает под собой не только физическую агрессию, но и психологическое давление, запугивание, распространение сплетен, вымогательство, издевательства.

Буллинг делится на прямые и непрямые формы агрессии. Физическая и вербальная формы агрессии являются прямыми формами буллинга, состоящими из явного выражения власти. Физический буллинг включает в себя любой тип физического нападения, например, удары, толчки, удары ногами, удушение и насильственное отнятие чего-либо у жертвы. Устные издевательства включают словесные домогательства или запугивания в форме обзывания, угроз, насмешек, злонамеренных насмешек и психологическое запугивание с использованием слов для причинения вреда жертвам. Воровство, вандализм, выражение лица или непристойные жесты также являются способами, с помощью которых дети подвергаются издевательствам. [28]

**3. МЕТОДЫ АВТОМАТИЧЕСКОГО ОБНАРУЖЕНИЯ ФАКТОВ ФИЗИЧЕСКОГО БУЛЛИНГА**

**3.1 Классификация методов выявления насилия (МВН)**

Быстрый рост объема видеоданных привел к возрастанию потребности в наблюдении и обнаружении аномалий. Такие аномальные события редко происходят по сравнению с нормальной деятельностью. Поэтому для уменьшения потерь труда и времени разработка автоматизированных систем видеонаблюдения для обнаружения аномалий стала насущной потребностью. Обнаружение аномалий в видео является сложной задачей, поскольку определение аномалии может быть неоднозначным и неопределенным.

Одним из важных аспектов обнаружения аномалий является распознавание и обнаружение насилия. Рост угроз безопасности во всем мире делает необходимым использование видеокамер для наблюдения за людьми, и тем самым раннее обнаружение и распознавание этих насильственных действий может значительно снизить эти риски. Методы моделирования, используемые для обнаружения аномалий или насильственных действий, могут быть широко классифицированы как неглубокие и глубокие модели. Основная цель нашей работы – это провести сравнительное исследование вышеперечисленных моделей.

Неглубокие методы моделирования – это методы, которые не способны самостоятельно изучать объекты, а скорее объекты, извлеченные с помощью ручных методов, должны быть предоставлены неглубокой сети для их классификации. Мелкая сеть может быть классификатором моделей, таких как машина опорных векторов (SVM), искусственная нейронная сеть (ANN) с одним скрытым слоем и т. д. Эти модели лучше всего подходят для обучения с подкреплением, которое должно быть хорошо маркировано. Основным недостатком этого метода моделирования является то, что они не адаптируются к изменениям паттерна автоматически. Кроме того, процесс маркировки может быть происходить вручную. Ллойд и др. [29] предложили дескриптор реального времени, который моделирует динамику толпы для обнаружения аномалий путем кодирования изменений в текстуре толпы с использованием временных сводок признаков матрицы встречаемости серого уровня, в которых k-кратная перекрестная валидация была выполнена для обучения классификатора случайного леса. Аналогично, Билински и Бремонд [30] использовали расширение улучшенных векторов Фишера (IFV), которое позволяет представлять видео с использованием как локальных признаков, так и их пространственно-временных позиций для распознавания и обнаружения насилия. Их результаты показали значительное улучшение в четырех общедоступных стандартных эталонных наборах данных.

В отличие от неглубоких моделей, большинство глубоких моделей не требуют отдельного экстрактора признаков, так как они основаны на методе изучения признаков, который заключается в том, что они изучают свои собственные признаки из заданных данных и классифицируют их на основе них. Кроме того, помимо сквозного обучения, вышеупомянутые извлеченные функции могут быть даны в качестве входных данных для SVMs и других классификаторов неглубоких моделей. Другой способ реализации глубоких моделей - это использование признаков из созданных вручную дескрипторов признаков и предоставление их глубокому классификатору. Эти модели работают как с контролируемыми, так и с неконтролируемыми методами обучения, но лучше подходят для последних. Несмотря на то, что они работают с немаркированными данными, они требуют больших объемов данных и вычислительной мощности. [31] предлагают сверточный пространственно-временной автоэнкодер для изучения регулярных паттернов в обучающих видеороликах для обнаружения аномалий. Несмотря на то, что модель может обнаруживать аномальные события и устойчива к шуму, в зависимости от того, насколько сложна деятельность, может возникнуть больше ложных выходов.

Методы обнаружения насилия (МОН) классифицируются на три категории в зависимости от используемого классификатора: МВН с использованием машинного обучения, МВН с использованием SVM и МВН с использованием глубокого обучения. SVM и глубокое обучение классифицируются отдельно, так как эти алгоритмы широко используются в компьютерном зрении. Подробная информация о каждом методе представлена в таблице. Методы представлены в хронологическом порядке. Также представлены методика обнаружения объектов и метод извлечения признаков. Общая система обнаружения аномалий или насильственных действий показана на Рисунке 1. Далее, в обзоре рассматриваются методы для автоматического обнаружения насилия, физического буллинга и агрессии.

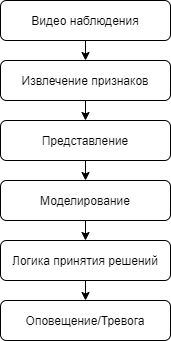


Рисунок 1. Общая схема системы обнаружения насилия

**3.2 Обнаружение насилия с использованием методов машинного обучения**

Насилие – это подозрительные события или действия в обычной жизни. Распознавание таких действий в видеозаписях наблюдения с помощью компьютерного зрения становится активной темой в области обнаружения действий [32]. Многие исследователи предлагали различные методики и методы выявления насильственных или ненормальных событий по мере быстрого роста уровня преступности для более точного выявления. Представлены различные методики выявления насилия, предложенные в последние годы. Методы обнаружения насилия классифицируются на три категории в зависимости от используемого классификатора: МОН с использованием машинного обучения, МОН с использованием SVM и МОН с использованием глубокого обучения. SVM и глубокое обучение классифицируются отдельно, так как эти алгоритмы широко используются в компьютерном зрении. Подробная информация о каждом методе представлена в таблицах. Методы представлены в хронологическом порядке. Также представлены методика обнаружения объектов и метод извлечения признаков.

Здесь обсуждаются методы обнаружения насилия, использующие в качестве классификатора различные традиционные алгоритмы машинного обучения, такие как KNN, Adaboost и другие.

В таблице 1 представлен перечень методик обнаружения, использующих различные методы классификации. Представлены также положения метода классификации, используемые в предлагаемой методике.

**Таблица 1. Методы выявления насилия с использованием различных методов классификации**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Метод | Метод Обнаружения Объекта | Метод Извлечения Признаков | Метод классификации | Тип Происшествия | Точность % |
| Движения больших двоичных объектов (AMV) ускорение векторная физическая величина, характеризующая способ для обнаружения быстрых изменений в видео [33] | Метод обнаружения эллипса | Алгоритм нахождения ускорения | Использование пространственно-временных признаков для классификации | Любой | 90% |
| Метод RIMOC фокусируется на скорости и направлении движения объекта на основе HOF (гистограмма оптического потока) [34] | Метод ковариационной матрицы на основе пространственно-временной векторный метод | Пространственно-временной векторный метод | SVM с помощью обучения с учителем | И многолюдно и менее многолюдно | 82 - 90 |
| Метод включает в себя двухэтапное обнаружение насилия и лиц в видео с использованием дескриптора VIF и алгоритмов нормализации [35] | VIF метод распознавания возражений CUDA и комплект детектора лица | Метод усадки рог для гистограммы | Интерполяционная классификация | Менее многолюдно | Lower frame rate 14% too high rate of 35 fs/s 97% |
| SVM-метод распознавания, основанный на статистической теории без декодирования видеокадров [36] | Метод векторной нормализации | Метод макроблоков для извлечения объектов | Движение области и дескриптор для классификации видео | Многолюдно | 96.1% |
| Обнаружение боев с движущимися кадрами [37] | Бинаризация изображений | Пространственно-временной метод извлечения сгустков | Классификация на основе длины блоба, наибольший рассмотрите боевой | Многолюдно | 70% - 98% |
| Лагранжевы поля направления и напрашивается словесная основа для признания насилия в видео [38] | Глобальная компенсация движения объекта | Лагранжева теория и ступенчатый метод извлечения признаков движения | Поздний сплав для классификации | Многолюдно | 91% - 94% |
| Простой подход к формированию видео для предварительной обработки с последующим извлечением признаков и распознаванием нормального и аномального события [39] | Гауссова Модель | Примените различные формулы к последовательному кадру, чтобы извлечь необходимый объект | Классификация на основе правил с использованием порога по умолчанию | Менее многолюдно | 90% |
| Совершенно другая техника, в которой учится для борьбы с животными обнаруживать человеческую борьбу [40] | Область движения и метод оптического потока | Методы Vif, Vif и IfV | Трансфертный подход к обучению | Менее многолюдно | 90% |
| Фреймворк для обнаружения насилия устанавливает параметр, который насильственные люди просто анализируют в социальных сетях [41] | Подход к демографическому анализу | Этническая структура | Модели машинного обучения | Менее многолюдно | 95% |

**1) Быстрое обнаружение драки**

В рамках компьютерного зрения распознавание действий становится актуальной темой. Для выявления последовательностей насилия предложен новый метод [33]. Предполагается, что в сценах драки движущиеся кадры имеют определенную форму и положение. Во-первых, для абсолютных изображений вычисляется разница между последовательными кадрами. Затем полученное изображение бинаризуется, приводя к числу движущихся скоплений и помечая самый большой из них в последовательности драки и на сцене без драки. Выбираются только K самых больших сгустков движения. Чтобы классифицировать k больших двоичных объектов, вычисляются различные параметры, такие как центроид, площадь, периметр и расстояние между большими двоичными объектами. Затем капли характеризуются как борьба и не-борьба. Эксперименты, проведены с использованием данных фильмов, уже около 200 клипов, набора хоккею, в которым есть 1000 роликов и набор данных УКУ-111, состоящий из настоящих роликов собранных с ютуба. Он имеет значительно более быстрое вычислительное время, что делает его приемлемым для приложений в режиме реального времени.

**2) Вращательно-инвариантный признак моделирования когерентности движения (RIMOC)**

Поскольку события агрессии трудно поддаются определению из-за отсутствия последовательности и часто нуждаются в интерпретации высокого уровня, именно поэтому принято классифицировать то, что часто присутствует в видеороликах с насильственным поведением людей на низком уровне, то есть неструктурированные и отрывистые движения. Для достижения этой цели предлагается инновационная задача-специфическое вращательно-инвариантное моделирование когерентности движения (RIMOC) [34]. Представленный метод был создан на собственных значениях, полученных из гистограммы векторов оптического потока (HOF) из моментов последовательного временного, плотно и локально вычисленного и, кроме того, встроенного в сферическое Риманово многообразие. Этот метод используется для изучения моделей статистики в еженедельном контролируемом режиме. Многомасштабная схема, применяемая на основе метода вывода, позволяющего воспринимать события с неустойчивым движением во времени и пространстве, как хорошие кандидаты на насильственные события. Нет набора данных, доступного специально для агрессивных событий.

**3) Быстрое распознавание лиц**

Для достижения цели обнаружения лиц в сценах насилия в помощь средствам контроля безопасности предлагается метод быстрого обнаружения лиц [35]. Дескриптор насильственного потока (ViF) используется вместе с Horn-Schunck для обнаружения сцен насилия на ранней стадии. Затем для улучшения качества видео применяется алгоритм неадаптивной интерполяции суперразрешения. Наконец, срабатывает детектор лица Kanade-Lucas-Tomasi (KLT). Чтобы достичь очень быстрой обработки, детектор лиц и алгоритм суперразрешения параллельны CUDA. CUDA состоит из функций, которые выполняются одновременно в многочисленных легких потоках на графическом процессоре. Эксперименты проводятся для оценки предложенного метода с использованием набора данных Boss, а другой набор данных называется violent dataset. Многообещающие результаты достигаются при обнаружении граней с точки зрения площади под кривой (AUC) и точности.

**4) Распознавание насильственной активности без декодирования**

Для целенаправленного обнаружения движений и их отслеживания в большинстве методов распознавания активности характерны сложность и ограниченность применения таких методов. Именно поэтому предлагается быстрый метод распознавания насильственной активности, основанный на векторах движения [36]. Во-первых, векторы движения извлекаются непосредственно из сжатых видеопоследовательностей. Во-вторых, атрибуты векторов движения анализируются в каждом кадре и между кадрами и достигают дескриптора вектора движения области (RMV). Наконец, радиальная основа была взята с использованием SVM в качестве функции ядра для классификации RMV и изучения того, присутствует ли активность насилия или нет. Существует множество наборов данных, доступных для распознавания активности, но ни один из них не подходит непосредственно, поскольку они сосредоточены на простом индивидуальном распознавании активности. Именно поэтому для оценки предложенного метода строится набор данных VVAR10, состоящий из 296 положительных и 277 отрицательных выборок. Образцы взяты из YouTube, UCF50, UCF sports и HMDB51. Эксперименты проводятся с использованием набора данных VVAR10, и результаты показывают, что он может обнаруживать 96,1% насильственных действий в видеопотоках, а скорость расчета быстра с точки зрения точности, частоты пропущенных сигналов тревоги (MAR) и частоты ложных сигналов тревоги (FAR), поэтому он подходит для встраиваемых систем.

**5) Автоматическое обнаружение драки**

Для обнаружения драк малозатратным и естественным способом предлагается подход, основанный на анализе движения [37]. Для обнаружения драки применяются два подхода. Первый - это двухуровневая статистическая агрегация, которая генерирует набор признаков. Пиксель движения извлекается, а затем области движения из серии кадров путем вычисления векторов оптического потока. Затем изображения оптического потока классифицируются по характеру векторов после устранения помех. После этого статистика движения вычисляется в соответствии с классифицированными типами для получения набора признаков для распознавания. Второй используемый подход - это сумка слов, которая используется для создания визуального набора слов. Затем гистограмма используется по всему набору визуальных слов в качестве вектора для характеристики видео для обнаружения драки. Эксперименты проводятся с использованием видеозаписей, основанных на реальных драках, и результаты эксперимента показывают, что предложенный метод превосходит существующие методы, основанные на дескрипторах MoSIFT с луковым механизмом и базовом анализе сигналов движения с луковым подходом с точки зрения отзыва и точности.

**6) Обнаружение множественной аномальной активности**

Системы видеонаблюдения теперь устанавливаются в торговых центрах, колледжах, школах, аэропортах и многих других местах из-за быстрого роста преступности. Поскольку видео снимаются 24/7, поэтому трудно контролировать их вручную и обнаруживать подозрительные действия, это увеличивает спрос на интеллектуальную систему наблюдения. Для решения этой задачи предлагается метод, который автоматически обнаруживает различные аномальные активности в видеоклипах [39]. Эта система включает в себя три основных этапа: обнаружение движущегося объекта, отслеживание объекта и понимание поведения для распознавания активности. На первом этапе предварительной обработки обнаруживаются движущиеся объекты и производится удаление шума. Затем процесс извлечения объектов используется для определения ключевых объектов, таких как направление, скорость, размеры и центроид. Извлеченные функции помогают отслеживать объекты в видеокадрах. На последнем этапе метод классификации на основе правил используется для классификации действий из видео, и если обнаруживается какая-то подозрительная активность, она генерирует сигнал тревоги. Эксперименты проводятся на вновь созданном наборе данных, основанном на 45 видео, которые содержат три вида деятельности: бег, ходьба и ползание. Результаты экспериментов показывают, что предложенный метод позволяет обнаруживать различные виды аномальной активности в различных сценариях и дает наилучшие результаты с точки зрения точности.

**7) Обнаружение насилия толпы**

В приложениях компьютерного зрения теория Лагранжа предоставляет богатый набор инструментов для анализа долгосрочной нелокальной информации о движении. На основе этой теории предложена специализированная лагранжева методика [38] для автоматического распознавания сцен насилия в видеопоследовательностях. Пространственно-временная модель на основе Лагранжевых полей направления используется для новых объектов и использует информацию о компенсации фонового движения, появлениях и долговременном движении. Расширенный подход применяется в позднем слиянии на основе каждого видео в качестве классификационной схемы для обеспечения соответствующего временного и пространственного масштаба объектов. Эксперименты проводятся для оценки предложенного метода с использованием трех эталонных наборов данных, которые представляют собой хоккейную драку, насильственную толпу и насилие в фильмах. Результаты показывают, что добавление теории Лагранжа является ценным признаком для обнаружения насилия, а эффективность классификации повышается по сравнению с современными методами, такими как ViF, HOG + BoW, two stream CNN и т. д. С точки зрения AUC и точности.

**8) Обнаружение межвидовой борьбы**

В социальной обработке сигналов обнаружение поведения людей в виде драк на видео имеет важное значение, особенно в контексте наблюдения. В реальной жизни сбор данных для обнаружения драк обычно ограничивает возможности машинного обучения и влияет на производительность современных методов, основанных на данных. Для решения этой проблемы представлен инновационный метод межвидового обучения наряду с набором малозатратных вычислительных характеристик движения для обнаружения драки [42]. Это позволяет эффективно избежать проблемы ограниченных данных о человеческой борьбе. Предложенный метод использует существенную общность между драками животных и человека, такую как физическое ускорение движущихся частей тела. Предлагаемый метод берет входные данные из видеозаписей борьбы животных и нескольких видеороликов борьбы людей. Предложен набор локальных признаков движения (LMF), основанный на статистике движения, сегментной корреляции, следующей парадигме анализа движения. LMF извлекаются из каждого видео. Временные характеристики, основанные на человеческой эвристике, извлекаются и для обнаружения драки принимаются традиционные алгоритмы машинного обучения, такие как SVM. Ансамблевые классификаторы предлагаются для выполнения межвидового обнаружения борьбы. Эксперименты проводились с использованием образцов видеоклипов, наборов данных хоккея и фильмов. Результаты экспериментов показывают, что предложенный метод работает лучше, чем современные методы, такие как ViF, OViF и сигналы движения с точки зрения точности.

**3.3 Методы обнаружения насилия с использованием SVM**

Здесь подробно обсуждаются методы обнаружения насилия, использующие в качестве классификатора машину опорных векторов (SVM).

В таблице 2 представлен перечень методов распознавания насильственных событий. SVM-это алгоритм, который используется для решения задач классификации с использованием обучения с подкреплением. В SVM мы строим данные на размерном пространстве (числовые объекты) и дифференцируем их в пределах двух классов. SVM-широко используемый метод в компьютерном зрении, так как он надежен и учитывает числовые характеристики. Он используется для задач, связанных с бинарной классификацией. SVM основан на ядре. Ядро - это функция, которая преобразует входные данные в высокомерное пространство, где решается задача. Основным недостатком SVM является недостаточная прозрачность результатов [43].

**Таблица 2. Методы обнаружения насилия с использованием SVM**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Метод | Метод Обнаружения Объекта | Метод Извлечения Признаков | Тип Происшествия | Точность % |
| Обнаружение насилия в переполненных сценах в режиме реального времени [44] | Дескриптор ViF | Набор функций | Многолюдный | 88% |
| Фреймворк Bag of words, использующий ускорение для обнаружения действий [45] | Алгоритмы фонового вычитания | Метод оценки эллипса для последовательных кадров | Менее многолюдно | Приблизительно 90% |
| Генетический алгоритм фреймворка с модулем отслеживания и обнаружения [46] | Гауссова модель | Алгоритм для извлечения оптического потока | Многолюдно | 82%-89% |
| Мультимодельные особенности фреймворка на базе подкласса [47] | Изображение CNN и ImageNet | Google Net для извлечения функций | Менее многолюдно | 98% |
| Определение частоты действий насильственного назначения [48] | Пространственные пирамиды и сетки для обнаружения объектов | Методы пространственно временной сетки для извлечения объектов | Многолюдно | 96%-99% с использованием различных наборов данных |
| Обнаружение насилия с использованием ориентированного насильственного потока [49] | Оптический метод потока | Сочетание ViF и OViF дескриптор | Многолюдно | 90% |
| AEI и HOG объединили фреймворк для распознавания аномального события в визуальных движениях [50] | Метод AEI для вычитания фона | HOG и пространственно - временные методы извлечения признаков | Как многолюдно, так и менее многолюдно | 94%-95% |
| Фреймворк включает в себя предварительную обработку, обнаружение активности и извлечение изображений. Эта работа идентифицирует аномальное событие и изображение из данных [51] | Оптический поток и временная разность для обнаружения объекта CBIR метод извлечения изображений | Функция Гаусса для анализа видеофайлов | Менее многолюдно | 97% |
| Поздний метод слияния для временных слоев восприятия для обнаружения активности высокого уровня. Использование несколько камер от 1 до N. [52] | Метод вектора движения для идентификации с нескольких камер в двух измерениях | SGT MtPL метод | Менее многолюдно | 98% |
| Двухканальная сверточная нейронная сеть для обнаружения в реальном времени [53] | ImageNet для обнаружения объекта | VGG-f модель для извлечения признаков | Многолюдно | 91%-94% |
| Решите проблему обнаружения, разделив цель по глубине и четкому формату с помощью Connect [54] | Обнаружение движения и модель TRof | BoW подход | Менее многолюдно | 96% |
| Метод мешка слов с использованием пространственно-временного метода для обнаружения аномалий в видео [55] | Представление сегментов и подсегментов | HOF и HOG для получения видеокадров | Менее многолюдно | 84%-91% |

**1) Обнаружение нарушения насилия в переполненных сценах в режиме реального времени**

Хотя использование камер видеонаблюдения популярно, но их эффективность вызывает сомнения. Для решения сложной задачи мониторинга насилия в местах массового скопления людей предлагается инновационный подход к обнаружению насилия в режиме реального времени [44]. Предлагаемый метод учитывает статистику изменения величины векторов потока во времени, рассматриваются две взаимосвязанные, но различные задачи: классификация насилия и выявление насилия. Основная цель состоит в том, чтобы обнаружить изменение от насильственного к ненасильственному поведению с наименьшей задержкой с момента, когда это изменение произошло. Бурные потоки (Vif) дескриптор используется для представления статистических данных, собираемых для нескольких последовательных кадров. Затем дескрипторы Vif классифицируются как насильственное или ненасильственное поведение с использованием линейного SVM. Эксперименты проводятся с использованием набора данных, основанного на видео наблюдения, собранных с YouTube посредством сбора данных с матчей по хоккею. Результаты экспериментов показывают, что предложенный метод работает лучше, чем существующие методы, такие как HOG, HOF и т.д., В зависимости от величины полей оптического потока только с точки зрения площади под кривой (AUC) и точности.

**2) Быстрое обнаружение насилия**

В компьютерном зрении проблема распознавания действий становится активной темой. Хорошо известный модуль bag-of-words(багаж слов) был использован в недавней работе, которая используется для распознавания борьбы. В этой схеме пространственно-временные признаки были взяты из видеокадров и использованы для классификации, а показатели точности 90% достигнуты для этой задачи. Вдохновленный результатами, которые говорят о том, что кинематические характеристики являются единственными дискриминантными для конкретных действий, инновационный метод, который использовал метод экстремального ускорения в качестве основного признака, предложил [45]. Чтобы эффективно оценить экстремальные характеристики, случайное преобразование применяется к последовательным кадрам видео. И использовал SVM и Adaboost в качестве классификатора. Были проведены эксперименты с использованием двух специальных наборов данных записей с хоккейных матчей и фильмов, и результаты показывают, что 12%-ное увеличение точности и работает лучше, чем современные универсальные методы распознавания действий, которые имеют такие функции, как масштабно-инвариантное преобразование признаков (SIFT) и масштабно-инвариантное преобразование признаков движения (MoSIFT) [56], а также предлагаемый метод в 15 раз быстрее.

**3) Гауссова модель оптического потока (GMOF)**

Для систем наблюдения обнаружение насилия становится актуальной темой, но изучается не так глубоко, как распознавание действий. Более ранние методы в основном фокусировались на обнаружении насилия и предпринимали мало усилий для определения места насилия. Вот почему была предложена надежная и быстрая система обнаружения и локализации насилия в местах наблюдения [46]. В рамках выделения областей-кандидатов насилия представлена Гауссовская модель оптического потока (GMOF). Модель адаптивно спроектирована как отклонение от нормального поведения толпы, обнаруженное в сценах. Затем для каждого видеообъема, который создается путем плотной выборки областей потенциального насилия, выполняется обнаружение насилия. Затем, чтобы отличить сцены насилия от ненасильственных сцен, был также предложен инновационный дескриптор под названием ориентационная гистограмма оптического потока (OHOF). Сначала обучающий модуль выбирает обучающие данные и извлекает дескриптор OHOF, а затем, используя линейный SVM, получает модель объекта. В конце, дескриптор сравнивается с обученной моделью SVM известных сцен насилия. Эксперименты проводятся с использованием трех сложных наборов данных. Результаты эксперимента показывают, что предлагаемый дескриптор работает лучше, чем предыдущие дескрипторы, такие как гистограммы ориентированных градиентов, гистограмма распределения оптического потока (Хоф), сочетание Hog и Хоф (ГНФ), MoSIFT и просеять в плане площади под кривой (AUC) и точность.

**4) Обнаружение насилия в видео с помощью подклассов**

Чтобы решить сложную проблему обнаружения насилия в видео, Ли и др. [47] специально фокусируется на сочетании мультимодальных функций путем добавления и использования подклассов, визуально связанных с насилием. Для достижения этой цели используется набор данных MediaEval 2015. Видео набора данных помечаются вручную по отношению к подклассам и приводят к 10 подклассам насилия, таким как кровь, пистолет, смерть и другие. Понятие подклассов взято из существующих исследований. SVM используется в качестве классификатора для обучения подклассов, и набор делится на два случайно непересекающихся набора, 70% для обучения и 30% для тестирования. Эта процедура также применяется к тестовому набору, и результаты показывают, что решение, основанное на подклассах, превосходит существующие методы, содержащие функции движения, такие как HOG, Motion Boundary Histogram (MBH) и HOF со средней точностью 0,303 и точностью 100 0,55 на средневековом наборе данных.

**5) Распознавание и выявление человеческого насилия**

Распознавание и выявление насилия становится важной темой для видеозаписей наружного наблюдения. Основная цель состоит в том, чтобы определить, происходит ли насилие. Во-первых, расширение улучшенных векторов Фишера (IFV) предлагается для видеоклипов [48]. Используются локальные объекты и их пространственно-временные положения, что позволяет репрезентировать видео. Затем для обнаружения насилия изучается популярный подход скользящего окна. Для ускорения подхода используется структура данных суммированной таблицы площадей и переформулируются формулы IFV. Во-вторых, локальные пространственно-временные характеристики извлекаются из видео с использованием улучшенных плотных траекторий (IDT). Затем видеопрезентация для каждого дескриптора вычисляется независимо, как HOG для представления видео с помощью IFV. Затем линейный SVM-классификатор используется для распознавания насилия, и в конце концов, используя подход быстрого скользящего окна, обнаруживается насилие. Обширная оценка проводится с использованием 4 современных наборов данных насильственных потоков, фильмов и хоккейных матчей. Набор данных Violence-Flow 21 используется для задачи обнаружения насилия. И результаты показывают, что предложенные подходы работают лучше по сравнению с существующими подходами, такими как ГНФ, Хоф.

**6) Обнаружение насилия с помощью ориентированного насильственного потока**

Спрос рынка на интеллектуальное обнаружение насилия постоянно растет с помощью камер видеонаблюдения, но все же это сложная область исследований. Во-первых, для практического обнаружения насилия в видео предлагается инновационный метод извлечения признаков, получивший название Oriented Violent Flows (OViF) [49]. В статистических ориентациях движения он в полной мере использует информацию об изменении величины движения. AdaBoost используется для выбора объектов, а затем SVM-классификатор обучается на выбранных объектах. Эксперименты проводятся на наборах данных базы данных Hockey и Violent-Flow для оценки полезности предложенного метода, и результаты показывают, что предложенный метод работает лучше, чем базовые методы: LTP и ViF с точки зрения точности и AUC. Во-вторых, принимаются стратегии комбинирования признаков и мультиклассификаторов, и достигаются выдающиеся результаты. Результаты эксперимента показывают, что использование комбинированных функций с AdaBoost и Linear-SVM позволяет повысить производительность по сравнению с современными методами на бенчмарке Violent-Flows.

**7) Автоматическая система видеонаблюдения в режиме реального времени**

В системе наблюдения обнаружение подозрительной деятельности играет жизненно важную роль. В среде ученых существует острая потребность в системе наблюдения, которая может надежно функционировать. Именно поэтому предлагается новая структура для системы видеонаблюдения в режиме реального времени, которая обнаруживает автоматически [51]. Работа разделена на три этапа. Фаза предварительной обработки включает в себя обнаружение аномальной активности человека и фазу извлечения изображений на основе контента. На этапе предварительной обработки все студенты должны зарегистрироваться перед началом курса обучения, а регистрация включает в себя сбор личных данных, и студент должен представить свою собственную фотографию для создания студенческого билета. Предложенная система нуждается в изображениях студентов также в различных условиях, таких как гнев, страх, печаль и т. д. получить точное описание в терминах поиска изображений на основе контента (CBIR). Эти записи сохраняются в базе данных CBIR для случая выявления аномальной активности студентов. Затем на следующем этапе изображение преобразуется в кадры. Алгоритм временной дифференцировки используется для обнаружения движущихся объектов, а затем с помощью функции Гаусса определяются области движения. Далее для распознанных объектов, которые являются человеческими или нечеловеческими, в качестве фильтра используется модель формы, основанная на уравнении омега. Человеческая деятельность классифицируется как нормальная и ненормальная деятельность с использованием SVM. В случае ненормальной деятельности человека, машина генерирует автоматическое предупреждение. Используя CBIR, он также вставляет метод извлечения распознанного объекта из базы данных для идентификации и распознавания объекта. Наконец, проведено программное моделирование с использованием MATLAB, и экспериментальные результаты показывают, что система достигает отслеживания, семантического обучения сцены и обнаружения аномалий в среде без участия человека.

**8) Надежное распознавание аномальной активности человека**

Для выявления любой ненормальной активности у пожилых людей и поддержки идеи качественной и независимой жизни предлагается надежная система распознавания ненормальной активности человека [50]. Вычисляя интеграцию векторов признаков, которые представляют собой гистограмму ориентированных градиентов (HOG) и моментов Цернике на средних энергетических изображениях (AEI), фреймворк структурируется для построения надежного вектора признаков. Компактное представление видеопоследовательностей обеспечивается формированием AEI без потери пространственно-временной информации. Взаимные размеры вектора признаков уменьшаются путем применения PCA, а для классификации видов деятельности применяется SVM. Предлагаемая работа оценивается на общедоступном 3D-наборе данных “Kinect Activity Recognition Dataset” (KARD) и “UR fall detection”. Эксперименты показывают, что результаты достигли 94% и 95,22% для набора данных UR fall и набора данных KARD с точки зрения средней точности распознавания (ARA) соответственно.

**9) Новая структура для анализа деятельности высокого уровня**

Предложена новая структура, основанная на позднем слиянии для анализа высокоуровневой активности с использованием мультинезависимых временных слоев восприятия [52]. Строятся два типа слоев восприятия, основанных на SVM и деревьях ситуационных графов (SGT). Структура состоит из трех этапов: многовременного анализа, многовременных слоев восприятия и позднего слияния. Результаты, полученные из слоев многовременного восприятия, сплавляются в счет активности на этапе позднего слияния. Для оценки этого подхода структура применяется к выявлению насильственных событий при визуальном наблюдении. Для экспериментов используются три известных набора данных: NUS-HGA, Behavior и некоторые видео с YouTube. Экспериментальные результаты показывают, что многовременная структура превосходит существующие одновременные структуры с точки зрения точности, показывая, что использование многовременного метода имеет преимущество перед одновременными методами.

**10) Обнаружение насилия в реальном времени**

Для надежного обнаружения действий насилия обычно недостаточно выбранных вручную функций. Поэтому модели на основе BI-каналы CNN и SVM для обнаружения насилия предлагается [53]. Предлагаемая модель состоит из трех частей: извлечение признаков, обучение SVM и слияние меток. Во-первых, структура Би-каналов CNN используется для извлечения двух особенностей. Первый признак-это исходный видеокадр, который используется для извлечения признаков внешнего вида, а второй вход-это разность соседних кадров, которая используется для разделения признаков движения. Затем внешний вид и движение принимаются в качестве классификатора для линейного SVM. В конце концов, результат обнаружения насилия достигается с помощью метода слияния меток, который объединяет информацию о движении и информацию о внешнем виде. Эксперименты проводятся для оценки предложенного метода с использованием наборов данных хоккейной борьбы и насильственной толпы. Результаты экспериментов показывают, что предложенный метод работает лучше, чем существующие методы типа HOG, HOF, MoSIFT, SIFT, Two-stream во многих реалистичных сценах с точки зрения точности.

**11) Breaking down violence**

Использование автоматических средств обнаружения насилия в видеозаписях имеет существенное значение для анализа камер видеонаблюдения и правоохранительных органов для поддержания общественной безопасности. Кроме того, это помогает защитить детей от попадания в неподходящий контент и помогает родителям принять лучшее решение о том, что их дети должны смотреть. Хотя это сложная проблема, так как определение насилия очень субъективно и широко. Вот почему обнаружение насилия в видеоклипах без присмотра человека является не только концептуальной, но и технической проблемой. Чтобы преодолеть эту проблему, предлагается идея насилия, используемая для CNN, разбивая ее на более конкретные и объективные части [54]. Во-первых, для изучения особенностей, непосредственно связанных с насилием, таких как кровь, взрывы, драки и т. д., используются независимые сети. Затем, чтобы описать насилие, используя такие признаки, отдельные классификаторы SVM обучаются для каждого понятия, а затем объединяют их позже в метаклассификацию. Также исследуется, как представить основанные на времени события в качестве сетевых входных данных для неподвижных изображений, поскольку многие изображения определяются в форме движения. Эксперименты проведены для оценки предложенного метода с использованием набора данных EvalMedia 2013 VSD. Результаты показывают, что концепция разбиения насилия на более мелкие концепции оказалась эффективным решением с точки зрения MAP@100 и AUC.

**12) Автоматическое определение стилей ведения драки**

Для классификации видеозаписей боевых искусств предложен метод распознавания [55]. На этапе предварительной обработки, чтобы очистить ненужные части видео, сначала обрезайте и сегментируйте соответствующий набор данных. Затем для извлечения объектов используются пространственно-временные точки интереса для обнаружения областей в видео, связанных с движением в последовательности кадров. Для классификации обучающий словарь всех дескрипторов строится и затем преобразуется в кластеры. Затем каждое видео представляется в виде гистограммы этих кластеров. Кроме того, KNN или SVM применяются в качестве классификатора к видео. Эксперименты проводились с использованием видеозаписей Олимпийских игр, включая Олимпийские игры 2012 года в Лондоне и юношеские Олимпийские игры 2014 года в Нанкине. То есть выбираются два основных класса видео: дзюдо и тхэквондо. Результаты экспериментов показывают, что предложенный метод с использованием SVM в качестве классификатора превосходит по точности существующие методы, использующие KNN и SVM с ядром RBF в качестве классификатора.

**3.4. Методы обнаружения насилия с использованием глубокого обучения**

Здесь подробно рассматриваются методы обнаружения насилия, использующие алгоритмы глубокого обучения. В таблице 3 представлен список методов распознавания, использующих классификацию на основе сверточных нейронных сетей (CNN, Convnet) [57]. Глубокое обучение основано на нейронных сетях. Этот метод используется для классификации насильственного распознавания на основе набора данных и извлеченных объектов с использованием сверточных слоев. Теперь методы обнаружения насилия, использующие алгоритмы глубокого обучения, подробно разрабатываются отдельно.

Таблица 3. Методы выявления насилия с использованием глубокого обучения

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Метод | Метод Обнаружения Объекта | Метод Извлечения Признаков | Тип Происшествия | Точность % |
| Обнаружение насилия с помощью 3D CNN [58] | 3D свертка используется для получения пространственной информации | Метод обратного распространения | Многолюдно | 91% |
| Глубокая архитектура для распознавания мест [59] | VGG VLAD метод поиска изображений | Метод обратного распространения для извлечения признаков | Многолюдно | 87%-96% |
| Жестокие сцены, используя СNN и глубоко-аудио особенности [60] | MFB | CNN модель | Многолюдно | Приблизительно 90% |
| Обнаружение жестоких видео с помощью convLSTM [61] | CNN вместе с ConvLSTM | CNN модель | Многолюдно | Приблизительно 97% |
| Выявление агрессивного поведения человека и с учетом траектории и глубоко СNN [62] | Глубокий CNN | Метод оптического потока | Многолюдно | 98% |
| Методология реорганизации леса Хафа [63] | Обнаружение объектов с использованием пространственно - временных признаков | Метод MoSIFT для извлечения видеофайлов | Менее многолюдно | 84%-96% |
| Обнаружение насилия с использованием пространственно-временных функций с помощью 3D CNN [64] | Нейронная сеть передает передает CNN полученную информацию | 3D CNN | Многолюдно | Приблизительно 97% |

**1) Обнаружение насилия с помощью 3D CNN**

Типичные методы обнаружения драк полагаются на знания предметной области для построения сложных объектов ручной работы из входных данных. С другой стороны, глубокие модели могут непосредственно действовать. Модель обучается с использованием контролируемого обучения, а градиентный спуск используется вместе с методом обратного распространения [58]. Эксперименты проводятся с использованием хоккейного набора данных, и результаты показывают, что предложенный метод работает лучше, не полагаясь на созданные вручную функции с точки зрения точности.

**2) Глубокая архитектура для распознавания места**

Для преодоления проблемы распознавания крупномасштабных визуальных мест, в которых основной задачей является быстрое и точное определение места по определенному запросу фотографий, была предложена архитектура CNN на основе слабо контролируемого распознавания мест [59]. Предлагаемый метод имеет три принципа. Во-первых, разрабатывается архитектура на основе CNN, которая обучается сквозным средствам для непосредственного выполнения задачи распознавания места. NetVLAD является ключевым элементом этой архитектуры, это новый всеобъемлющий слой VLAD, который поощряется локально агрегированными векторами дескрипторов, в основном используемыми для извлечения изображений. VLAD layer подходит для обучения с использованием обратного распространения и охотно подключается к любой архитектуре, основанной на CNN. Во-вторых, чтобы получить параметры архитектуры в сквозном средстве из изображений, представляющих одни и те же места в течение всего времени, которое берется из Google Street View Time Machine, разрабатывается недавно разработанная слабо контролируемая процедура обучения ранжированию на основе потерь. Наконец, эксперименты проводятся на основе типичного процесса оценки распознавания мест с использованием наборов данных Питтсбурга и Токио 24/7 в свободном доступе, и результаты показывают, что предлагаемая архитектура работает лучше, чем ненаучные представления изображений и готовые дескрипторы CNN на двух сложных тестах распознавания мест и современном представлении изображений на тесте поиска изображений.

**3) Обнаружение жестоких сцен, используя СNN и глубоких нейронных сетей**

Предложена система обнаружения насильственных сцен [60], использующая CNN, построенную на акустической информации из видеоклипов. Телеканал CNN подает информацию о звуке в двумя образами: в качестве классификатора и как экстрактор. Во-первых, 40-мерный Mel Filter-Bank (MFB) используется в качестве входного объекта для CNN с их Дельта-и. Затем видео преобразуется в короткие куски. Функции MFB разделены на 3 канала функций для изучения местных особенностей. После этого CNN используется для представления объектов. Функции на основе CNN используются для построения SVM-классификаторов. Затем обнаружение сцены насилия выполняется на каждом фрагменте видео. Далее обнаружение производится максимальным или минимальным объединением на уровне сегментов обнаружений. Эксперименты проводятся с помощью датасета MediaEval, и результаты показывают, что предлагаемый метод работает лучше, чем базовые методы: только аудио, только визуальный и аудио-выученный слияние и визуальный с точки зрения средней точности.

**4) Обнаружение жестоких видео с помощью сверточной долговременной кратковременной памяти (ConvLSTM)**

Для распознавания насилия в видео предложен метод на основе глубоких нейронных сетей [61]. CNN используется для извлечения функций из уровня кадра в видео. Затем эти функции накапливаются с помощью варианта LSTM, который использует сверточные ворота. Комбинация CNN и ConvLSTM может принимать локализованные пространственно-временные характеристики, что позволяет проводить локальный анализ движения, происходящего в видео. Также предлагается использовать разности соседних кадров в качестве входных данных для модели, которая кодирует изменения, произошедшие в видео. Эксперименты проводятся с использованием трех популярных наборов данных: хоккей, фильмы и видео насильстввенного характера. Результаты показывают, что предложенная модель работает лучше, чем современные методы, такие как ViF+Vif, ViF, three streams + LSTM и другие с точки зрения точности.

**5) Обнаружение насильственного поведения человека путем интеграции траектории и глубокого CNN**

Типичные методы обнаружения насилия обычно зависят от особенностей ручной работы, которая в основном не подходит. Вдохновленный работой глубоких моделей для распознавания человеческих действий, предложен инновационный метод обнаружения насильственного поведения человека путем объединения траектории и глубокого CNN [62], который использует преимущества как созданных вручную функций, так и глубоко изученных функций. Эксперименты проводятся на двух реальных наборах данных: хоккейная драка и насилие толпы. Полученные результаты показывают, что предложенный метод работает лучше, чем существующие методы: HOG, HOF, ViF и другие с точки зрения точности.

**6) Метод распознавания драки**

В компьютерном зрении распознавание действий становится важным направлением исследований. Такие задачи, как агрессивное поведение или драки, сравнительно мало изучены, но может быть полезно во многих сценариях видеонаблюдения, таких как тюрьмы, психиатрические палаты или личные смартфоны. Их обширная практичность создает интерес к разработке детекторов насилия или борьбы. Основным аспектом детекторов является эффективность, то есть эти подходы должны быть быстрыми в вычислительном отношении. Созданные вручную пространственно-временные характеристики достигают высокой точности как для внешнего вида, так и для движения, но извлечение некоторых характеристик все еще невозможно для реального применения. Впервые парадигма глубокого обучения применяется к задаче, использующей 3D CNN, которая принимает полную видеопоследовательность в качестве входных данных. Но особенности движения человека для этой задачи имеют решающее значение, и использование полного видео в качестве входного сигнала вызывает шум и избыточность в процессе обучения. Для этой цели был предложен гибридный фреймворк “handcrafted/learned” [63]. Метод, во-первых, направлен на получение иллюстративного изображения из видеопоследовательности, взятой в качестве входных данных для извлечения признаков, а лес Хафа используется в качестве классификатора. Затем, чтобы классифицировать это изображение и получить заключение для последовательности, используется 2D CNN. Проведены эксперименты с использованием трех хорошо известных наборов данных записей игр по хоккею, фильмов и сцен насилия. И результаты показывают, что предложенный метод работает лучше, чем различные методы ручного и глубокого обучения, основанные на точности и стандартных отклонениях.

**7) Обнаружение насилия с использованием пространственно-временных функций с помощью 3D CNN**

Для признания насильственных действий в целях безопасности обязательна усиленная система наблюдения, позволяющая избежать социального, экономического и экологического ущерба. Для этой цели предлагается структура трехэтапного сквозного обнаружения насилия в глубоком обучении [64]. Во-первых, в потоках видеонаблюдения люди обнаруживаются с помощью облегченной модели CNN, чтобы преодолеть и уменьшить огромную обработку непригодных кадров. Во-вторых, порядка 16 кадров с обнаруженными особями передаются в 3D CNN, где пространственно-временные характеристики этих последовательностей извлекаются и подаются в Классификатор Softmax. Затем 3D-модель CNN оптимизируется с помощью инструментария оптимизации нейронных сетей и открытого визуального вывода, разработанного Intel. Обученная модель преобразуется в промежуточную иллюстрацию и изменяет ее для исполнения на конечной платформе для окончательного обнаружения насилия. После обнаружения насилия сигнал тревоги передается в соседний отдел безопасности или полицейский участок, чтобы произвести действие. Эксперименты проводятся с использованием датасета драк толпы, хоккейных матчей и насилия в фильмах. Результаты эксперимента показывают, что предложенный метод работает лучше, чем современные методы, такие как ViF, AdaBoost, SVM, Hough Forest и 2D CNN, sHOT и другие с точки зрения точности и AUC.

**4. ОБЗОР ПРОГРАММНЫХ РЕШЕНИЙ ПО РАСПОЗНАВАНИЮ И КЛАССИФИКАЦИИ ДЕЙСТВИЙ НА ВИДЕО**

Действие тела также является критическим интерфейсом, привлекая все больше и больше внимания в последние годы. В этом разделе мы обобщаем исследования, связанные с действием тела, и обсуждаем следующие направления: распознавание действий, прогнозирование действий, непрерывный анализ действий, безопасность и анализ эгоцентрического видения действий.

Визуальное распознавание действий является основным методом в распознавании буллинга. С помощью записи траекторий [65] отслеживали монокулярные и изученные траектории движения человека для распознавания действий. С развитием глубокого обучения были разработаны различные глубокие сети для получения высокой производительности распознавания для приложений HRI. [66] понимали намерения человека с помощью двух контролируемых рекуррентных нейронных сетей множественных временных масштабов (MTRNN). Первый слой распознает человеческие действия, а второй слой предсказывает человеческие намерения, основываясь на результатах первого слоя. Кроме того, [67] изучили особенности действия иерархическим образом, чтобы понять спонтанные эмоции для естественного эмоционального взаимодействия между человеком и роботами. Этот подход применил многоканальную сверточную нейронную сеть (CCNN) для интеграции мультимодальных функций распознавания невербальных эмоций, включая мимику и движения тела.

Информация о глубине полезна для описания действий человека в трехмерных пространствах. Как и традиционные подходы на основе RGB-видео, траектория также эффективна для описания действий с помощью 3D-информации о действиях. [68] пересекли суставы тела в направлениях и записали траектории этих направлений в последовательности действий. Они моделировали траектории и представляли активность в виде гистограммы векторов направления для интерактивного и неинтерактивного распознавания действий. Поскольку легко обнаружить суставы и части тела с захваченными скелетами, были предложены подходы к изучению и присвоению Весов важности суставам тела для эффективного представления действий [69]. Piyathilaka et al. [70] включили веса важности для 3D-скелетных суставов, чтобы игнорировать запутанные или нерелевантные функции, и обучили динамические байесовские сети (DBN) для классификации человеческой деятельности, чтобы еще больше улучшить возможности домашних сервисных роботов. Для представления действия были разработаны различные эффективные совместные функции. [71] представили Контрастивную модель распределения признаков (CFDM) в человеческих взаимодействиях, которая ввела внутрикадровый скелет и изучила дискриминационное представление для распознавания взаимодействия человека и человека. Кроме того, слияние признаков представлено для решения задачи распознавания сложных действий [72]. Charalampous et al. [73] позволили роботу ориентироваться в населенной среде, чтобы продемонстрировать социально приемлемое поведение. Он распознал человеческие действия с помощью RGB-видео и последовательностей глубины. Комбинируя информацию о глубине с информацией RGB, он способен повысить точность распознавания действий и способствовать применению действий в распознавании буллинга и насилия.

Взаимодействие человека и объекта-еще одно часто обсуждаемое направление. Поскольку она включает в себя человеческие действия и объекты, графовая модель, естественно, используется для представления взаимодействия. Коппула и др. [74] приняли структуру графа для моделирования взаимодействий человека и объекта, где узлы представляли объекты и под-действия, а ребра представляли отношения между объектами, отношения между объектами и под-действиями и их эволюцию во времени. Наконец, была представлена структурная векторная машина поддержки (SSVM) для распознавания взаимодействий. Кроме того, иерархическая обработка также представлена для решения этой проблемы. Granata et al. [75] представили двухслойный подход, состоящий из слоя восприятия и слоя классификации, для распознавания человеческой деятельности. На уровне восприятия траектории суставов скелета использовались для представления виртуальной динамики человека. Классификация слоя используется многоступенчатая модель опорных векторов (MSVM) признать объект участвует вспомогательных действий. Дискретный скрытый Марков был использован для распознавания взаимодействия. Кроме того, Dutta et al. [76] использовали вероятностную машину состояний (PSM) для моделирования человеческих поз, объектов и их отношений для распознавания действий. Эти подходы моделировали человеческие действия, объекты и их отношения для распознавания взаимодействия, имитируя реальные взаимодействия в HRI. Поэтому они обеспечили теоретическую подготовку к применению HRI.

**5. ОБЗОР СУЩЕСТВУЮЩИХ НАБОРОВ ДАННЫХ ДЛЯ РАСПОЗНАВАНИЯ И КЛАССИФИКАЦИИ ДЕЙСТВИЙ НА ВИДЕО**

В настоящее время существует несколько десятков наборов данных, которые посвящены различным аспектам распознавания тех или иных данных с видео. Готового набора данных с размеченными действиями, которые можно считать проявлениями буллинга или агрессивного поведения среди школьников, в открытом доступе обнаружено не было. Однако среди доступных наборов данных есть такие, которые содержат классы действий напоминающие агрессивное поведение. Ниже представлен перечень доступных наборов данных с описанием, которые будут учтены при формировании набора данных для решения поставленных в проекте задач. Перечислены следующие наборы данных:

**HACS (Human Action Clips and Segments Dataset for Recognition and Temporal Localization** - Набор коротких видео и сегментов, содержащих человеческие действия, предназначенный для распознавания и временной локализации) [77-78].

Краткое описание: Набор данных содержит два вида ручной разметки: «HACS Clips» состоит из 1,55 миллионов размеченных 2-секундных клипов из 504 тысяч видео; «HACS Segments» описывает сегменты, содержащие завершенные действия (от начала действия до его окончания) в 50 тысячах видео. Предполагаемое назначение набора данных - предобучение моделей распознавания и локализации действий (набор данных HACS включает в себя набор данных SLAC).

**Moments in Time Dataset (A large-scale dataset for recognizing and understanding action in videos - Большой набор данных для распознавания и понимания действий в видео).** Набор данных состоит из миллиона размеченных 3-секундных видео с людьми, животными, объектами и природными явлениями, содержащих динамические сцены [79-80].

**AVA-Kinetics Dataset.** Объединение наборов данных AVA Actions dataset (430 видео длительностью 15 минут, в которых аннотирована каждая секунда) и Kinetics dataset (650000 видео клипов длительностью 10 секунд с описанием действий, совершаемых людьми) [81-82].

**Oops! A Dataset of Unintentional Action.** Набор данных состоит из 20723 видео (более 50 часов) компиляций неудачных ситуаций из YouTube. Видео сняты любителями и содержат различные действия, среды и ситуации из реальной жизни: физические и социально-бытовые ошибки, ситуации с недостаточными физическими навыками или связанные с реакцией на различные факторы окружающей среды [83-84].

**20BN-something-something Dataset.** Краткое описание: Набор данных содержит 220847 видео с 318572 аннотациями. На видео представлены люди, совершающие заранее определенные действия с повседневными объектами. Среди действий имеются: толкание предметов, броски предметов, удары предметами [85-86].

**ActivityNet** (A Large-Scale Video Benchmark for Human Activity Understanding - Большой проверочный набор данных по пониманию действий людей). Набор данных содержит около 20000 видео с разметкой 200 видов деятельности людей [87-88].

**Hockey Fight dataset – Набор сцен драк с хоккейных матчей**

Этот набор данных был представлен Nievas et al. [89] и содержит 1000 коротких видеоклипов, взятых из Национальной хоккейной лиги (НХЛ). В этом наборе данных 500 видеоклипов помечены как боевые, а 500-как не боевые. Каждый клип состоит из 50 кадров с разрешением 360 × 288 пикселей. В классе fight все клипы связаны с боями на хоккейных площадках, а класс non-fight также связан с той же средой, содержащей клип non-fight, чтобы надежно обнаруживать сцены насилия в спортивных видео. Некоторые примеры кадров из этого набора данных приведены на Рисунке 2.



**Рисунок 2.** Примеры видеокадров, случайно выбранных из: (а) жестокой толпы, (б) насилия в кино и (в) хоккейной драки.



**Рисунок 2.** Примеры видеокадров, случайно выбранных из: (а) жестокой толпы, (б) насилия в кино и (в) хоккейной драки.

**BEHAVE dataset.** Снятые статической камерой пять видео (разрешение изображения 640 × 480 пикселей) с более чем 5000 кадрами составляют набор данных, включающий групповое обсуждение, ходьбу, бег, погоню и драку, с нарушением езды на велосипеде и проезжающих автомобилей. Поскольку в этом наборе данных только 392 кадра помечены как “ненормальное взаимодействие”, каждый эпизод насильственного характера содержит 19 кадров. 80 мини видео без сцен с драками случайным образом выбираются из кадров, помеченных как “нормальное взаимодействие”, причем 19 кадров также включены в каждый отрезок(эпизод) [90].

**Crowd Violence dataset.** Этот набор данных в основном представляет поведение насилия толпы, и большинство сцен являются динамическими, что значительно увеличивает сложность обнаружения. На этом наборе данных собрано 246 видеоклипов (разрешение изображения 320 × 240 пикселей) с 123 насильственными выборками и 123 ненасильственными выборками [91].

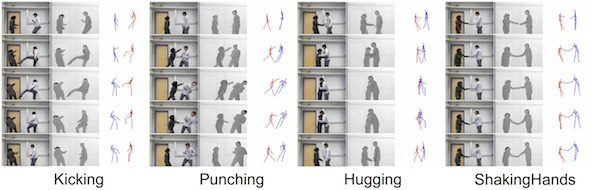
**Violent Crowd.** Набор данных о групповом насилии толпе был представлен Hassner et al. [92]. Этот набор данных содержит 246 видео, взятых с YouTube, представляющих различные типы сцен и сценариев. Сначала набор данных содержит пять наборов видеоклипов. В каждом наборе есть две категории: насильственные и ненасильственные. Для экспериментов мы объединили эти пять наборов, чтобы сформировать две категории, где 123 видеоклипа связаны с насильственными событиями, а 123 видео связаны с ненасильственными клипами. Каждый видеоклип имеет разрешение 320 × 240 пикселей с длиной от 50 до 150 кадров.

**Violence in Movies.** Этот набор данных был введен Nievas et al. [93] для обнаружения драк, и он состоит из 200 видеоклипов, в которых видео драк людей были взяты из действий с фильмов, в то время как видео без боя были извлечены из общедоступных наборов данных для распознавания действий. Этот набор данных охватывает множество сцен со средним разрешением 360 × 250 пикселей, и каждый клип ограничен 50 кадрами. В этом наборе данных первый человек в последовательности имеет низкое или полное отсутствие движения камеры. Некоторые примеры кадров из этого набора данных приведены на рис.6.

**Violent-Flows.** Набор данных Violent-Flows (Crowd Violence \ Non-violence) [94] содержит 246 видеороликов с реальными кадрами насилия толпы. Видео собираются с YouTube и содержат различные сцены, например, улицы, футбольные стадионы, волейбольные и хоккейные арены, а также школы. Набор данных разделен на 5 складок, и мы следуем рекомендуемой 5 - кратной перекрестной проверке, чтобы сообщить о производительности.

**MediaEval-2013-VSD dataset.** MediaEval-2013-VSD data set [95] - это эталонный набор данных о сценах насилия, который содержит 25 голливудских фильмов различных жанров. Набор данных обеспечивает сегментацию кадров из фильмов, где каждый сегмент был вручную аннотирован, чтобы отличить, имело ли место физическое насилие в сценах для каждого фильма. Определение насилия, используемое в конкурсе, заключается в том, что сцена является насильственной, если “никто не позволит восьмилетнему ребенку увидеть”. Набор данных, выпущенный с уже разделенными разделами в обучающих и тестовых наборах. Обучающий набор включает в себя 18 фильмов, в то время как тестовый набор состоит из 7 фильмов Примечательно, что среди всех сцен только 20% из них были классифицированы как насильственные. Хотя набор данных содержит аннотации для отдельных концепций (например, кровь, драки и т. д.), эти аннотации доступны только для обучающего набора.

**SBU Kinect Interaction.** Восемь типов взаимодействия двух человек были собраны с помощью сенсора Microsoft Kinect (3,3 ГБ). Авторы собирают восемь взаимодействий: приближение, уход, толчки, удары ногами, удары кулаками, обмен предметами, объятия и рукопожатия от семи участников и 21 пары наборов из двух актеров. Каждый кадр содержит цветное изображение, карту глубины и 3-мерные координаты 15 суставов от каждого человека [96].



**Caviar.** Команда CAVIAR собрала видеопоследовательности ground truth для оценки эффективности проекта, а также для использования сообществом компьютерного зрения. Основная истина доступна в интернете и закодирована в XML-варианте, разработанном нами для использования с данными компьютерного зрения [97].

Всего существует 52 последовательности, состоящие из 90 тысяч кадров. Около 1/3 - это крытый офисный вестибюль. Другие 2/3 находятся в торговом центре, откуда открывается 2 вида на ту же самую деятельность.

Сценарные (и ненаписанные) действия включают в себя: прогулка, просмотр, спад, левый объект, встреча, драка, витрина магазина, вход в магазин, выход из магазина.

Наземная истинность движущихся целей была составлена вручную и включает в себя информацию о геометрическом положении, кратковременные описания активности и долгосрочные описания активности. Маркировка groundtruth для некоторых людей (108) в некоторых видеопоследовательностях (19) была расширена, чтобы также отметить голову, руки, ноги и плечи. Всего было аннотировано 52+K целевых кадров. Эта дополнительная маркировка не использовалась в проекте, но была объявлена и предоставлена для использования международным сообществом.

Одна последовательность с 3 маркировками была использована для экспериментов по консистенции маркировки. Результаты показали хорошую геометрическую и временную согласованность. Но статус семантической маркировки сбивал с толку: возможно, 1% ошибок маркировки и 40+% семантической онтологии/временных различий.

Здесь показан пример изображения из наземной правды с аннотированными людьми и головами и т. д.



**KARD.** Публичный набор данных распознавания активности Kinect (KARD) [98-99], который содержит 18 действий, разделенных на десять жестов и восемь действий, каждое из которых выполняется три раза десятью различными субъектами. Третий вклад заключается в валидации предложенного метода по отношению к хорошо известному общедоступному набору данных.

Авторы собрали набор данных, названный KARD, уделяя особое внимание правильности как самих полученных данных, так и основной правды. Кард содержит 18 действий, разделенных на десять жестов (горизонтальный взмах руки, высокий взмах руки, два взмаха руки, высокий бросок, нарисуйте x, нарисуйте ТИК, удар вперед, боковой удар, изгиб и хлопок в ладоши) и восемь действий (поймайте колпачок, бросьте бумагу, возьмите зонтик, прогуляйтесь, позвоните по телефону, выпейте, сядьте и встаньте).

Различие между этими двумя классами действий полезно для лучшей оценки эффективности системы как на простых последовательностях, которые отдельно включают определенные части тела, то есть жесты, так и на сложных действиях, где различные части тела взаимодействуют друг с другом.

Каждое упражнение было повторено трижды десятью различными индивидуумами (девять мужчин и одна женщина) в возрасте от 20 до 30 лет и ростом от 150 до 185 см. Пользователям были даны инструкции о том, какое действие выполнять, например, “хлопать в ладоши”, “ловить шапку”, без предоставления информации о том, как это делать, чтобы гарантировать естественность движений.

Набор данных был захвачен с помощью устройства Kinect, расположенного примерно в 2-3 м от объекта исследования, в офисной сцене, содержащей письменный стол, телефон, вешалку для одежды и мусорное ведро.

KARD состоит из 540 последовательностей примерно за 1 час видео, снятых с разрешением 640 × 480 пикселей со скоростью 30 кадров в секунду. Для каждой последовательности мы предоставляем как изображения RGB и глубины, так и список обнаруженных стыков в реальном мире и экранных координатах.

**Media Eval.** Набор данных, а именно LIRIS-ACCESSIVE [100-102], содержит 9800 видео отрывков, совместно используемых под лицензиями creative commons, что позволяет выпустить базу данных без проблем с авторскими правами. Описаны фильмы, из которых были извлечены выдержки, и продемонстрировано разнообразие базы данных. Мы также представляем новый экспериментальный протокол ранжирования видео-фрагментов базы данных по оси индуцированного возбуждения. Кроме того, все отрывки в ЛИРИСЕ ранжируются с помощью краудсорсинга в широко используемом двумерном пространстве валентности-возбуждения. Это многомерное пространство было предпочтено другим категориальным подходам, которые классифицируют эмоции на небольшое число дискретных кластеров и могут не отражать сложность, разнообразие и богатство эмоций, которые могут быть вызваны таким огромным количеством видео.

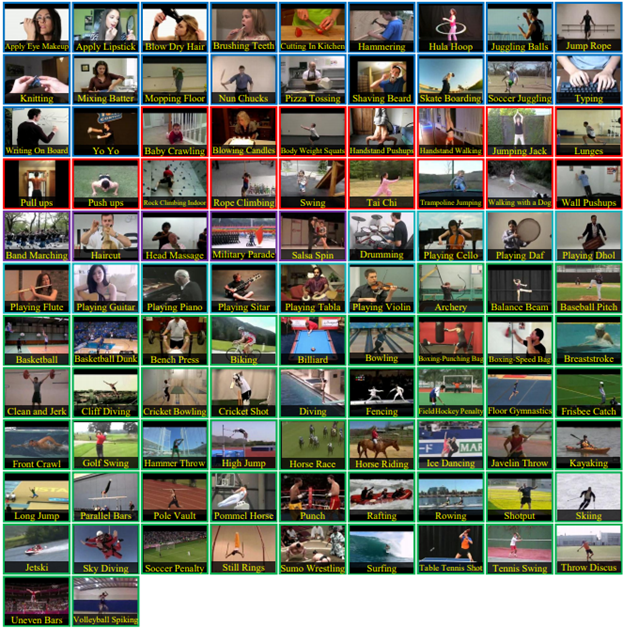
Лирис-присоединиться состоит из шести коллекций:

* ***Дискретные рейтинги валентности и возбуждения, для 9800 коротких видео отрывков, извлеченных из 160 фильмов.*** Оценочные аффективные оценки также доступны.
* Непрерывный Лирис-присоединение - непрерывная индуцированная валентность и самооценка возбуждения для 30 фильмов. Также доступны необработанные и постобработанные измерения GSR.
* ***Средневековые 2015 аффективное воздействие фильмов загрузки задач*** - аннотации насилия и аффективного классов для 9800 выдержки из дискретных Лирис-удовлетворить частично, плюс дополнительные 1100 выдержки используются, чтобы продлить испытательный комплект для средневекового 2015 аффективное воздействие задача кино.
* ***Средневековый 2016 эмоциональное воздействие фильмов загрузки задач*** - тест набор для средневекового 2016 эмоциональное воздействие на целевую фильмы: 1200 дополнительные видео отрывки на глобальные подзадачи аннотации и еще 10 фильмов для непрерывной подзадачи аннотации.
* ***Средневековый 2017 эмоциональное воздействие фильмов загрузки задач*** - Валанс/возбуждение и страх аннотации для разработки и тестирования задает средневекового 2017 эмоциональное воздействие задача кино. Кроме того, предоставляются визуальные и звуковые функции.
* ***Средневековый 2018 эмоциональное воздействие фильмов загрузки задач*** - Валанс/возбуждение и страх аннотации для разработки и тестирования задает средневекового 2018 эмоциональное воздействие задача кино. Кроме того, предоставляются визуальные и звуковые функции.

**UCF 101.** Набор данных состоит из веб-видео, которые записываются в неограниченной среде и обычно включают движение камеры, различные условия освещения, частичную окклюзию, кадры низкого качества и т. д. На рисунке ниже показаны примеры фреймов классов действий из UCF101 [103-104].

Классы действий: UCF101 включает в себя общее количество 101 класса действий, которые мы разделили на пять типов: взаимодействие человека и объекта, только движение тела, взаимодействие человека и человека, игра на музыкальных инструментах, Спорт.

UCF101-это расширение UCF50, которое включает в себя следующие 50 классов действий: {бейсбольное поле, стрельба по баскетболу, жим лежа, езда на велосипеде, бильярдный выстрел, Брасс, толчок, дайвинг, барабанная дробь, фехтование, гольф-качели, Прыжки в высоту, скачки, верховая езда, хула-хуп, метание копья, жонглирование мячами, Прыжки домкратом, скакалка, каякинг, выпады, военный парад, смешивание теста, монашеские патроны, бросание пиццы, игра на гитаре, игра на фортепиано, игра на табле, игра на скрипке, прыжки с шестом, скачки на луке, подтягивания, удар кулаком, отжимания, скалолазание в помещении, скалолазание по канату, Гребля, сальса-спины, скейтбординг, Лыжи, Скиджет, футбольное жонглирование, качели, Тайчи, теннисные качели, метание диска, прыжки на батуте, волейбол с шипами, прогулка с собакой, Йо-Йо}. Метки классов цветов указывают, к какому предопределенному типу действий они относятся.



**6. ЗАКЛЮЧЕНИЕ**

В ходе проделанной работы был сделан анализ существующих алгоритмов машинного обучения и ИИ для автоматического выявления и классификации физического, социального и другого рода буллинга, в интернет-пространстве. Был проделан сравнительный анализ этих алгоритмов на примерах других работ и исследований зарубежных специалистов.

В результате выполненных исследований по данному проекту за 2020 год получены следующие результаты научно-технической деятельности (РНТД):

Кроме всего этого, были опробированы множество дата сетов, каждая из которых описывалась выше и приводился пример его использования. На каждом дата сете были протестированы несколько алгоритмов. Это дало нам понимание какой из алгоритмов лучше использовать в рамках нашей исследовательской деятельности. На основе этого анализа были выведены критерии и была построена схема построения собственного дата сета видеозаписей со школьных камер видеонаблюдения.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

[1] Alsaker, F. D., & Nagele, C. (2008). Bullying in kindergarten and prevention. In W. Craig & D. Pepler (Eds.), An international perspective on understanding and addressing bullying (pp. 230–252). Kingston, Indiana, USA: PREVNet.

[2] Salmivalli, C., Kärnä, A., & Poskiparta, E. (2010). From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), Handbook of bullying in schools: An international perspective (pp. 441-454). New York, NY, US: Routledge/Taylor & Francis Group.

[3] Olweus, D. (1993). Bullying at school: What we know and what we can do. Malden, MA: Blackwell Publishing

[4] Ttofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011). The predictive efficiency of school bullying versus later offending: A systematic/meta-analytic review of longitudinal studies. Criminal Behaviour and Mental Health, 21, 80–89. doi:10.1002/cbm.v21.2

[5] Olweus, D. (1993). Bullying at school: What we know and what we can do. Malden, MA: Blackwell Publishing.

[6] Limber, S. P. (2011). Development, evaluation, and future directions of the Olweus bullying programme. Journal of School Violence, 10(1), 71-87. doi:10.1080/15388220.2010.519375.

[7] Olweus, D., & Limber, S. P. (2010). Bullying in school: Evaluation and dissemination of the Olweus bullying prevention program. American Journal of Orthopsychiatry, 80(1), 124–134. doi:10.1111/j.1939-0025.2010.01015.x

[8] Alsaker, F. D., & Nagele, C. (2008). Bullying in kindergarten and prevention. In W. Craig & D. Pepler (Eds.), An international perspective on understanding and addressing bullying (pp. 230–252). Kingston, Indiana, USA: PREVNet.

[9] Salmivalli, C., Kärnä, A., & Poskiparta, E. (2010). From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), Handbook of bullying in schools: An international perspective (pp. 441-454). New York, NY, US: Routledge/Taylor & Francis Group.

[10] Raskauskas, J. L., Gregory, J., Harvey, S. T., Rifshana, F., & Evans, I. M. (2010). Bullying among primary school children in New Zealand: Relationships with prosocial behaviour and classroom climate. Educational Research, 52(1), 1–13. doi:10.1080/00131881003588097

[11] MacIntyre, D., & Lawlor, M. (2016). The stay safe program. Dublin, Ireland: Department of Education and Skills.

[12] Alsaker, F. D., & Nagele, C. (2008). Bullying in kindergarten and prevention. In W. Craig & D. Pepler (Eds.), An international perspective on understanding and addressing bullying (pp. 230–252). Kingston, Indiana, USA: PREVNet.

[13] Raskauskas, J. L., Gregory, J., Harvey, S. T., Rifshana, F., & Evans, I. M. (2010). Bullying among primary school children in New Zealand: Relationships with prosocial behaviour and classroom climate. Educational Research, 52(1), 1–13. doi:10.1080/00131881003588097

[14] Biggs, B. K., Vernberg, E. M., Twemlow, S. W., Fonagy, P., & Dill, E. J. (2008). Teacher adherence and its relations to teacher attitudes and student outcomes in an elementary school-based violence prevention program. School Psychology Review, 37(4), 533–549.

[15] Brendgen, M., & Poulin, F. (2017). Continued bullying victimization from childhood to young adulthood: A longitudinal study of mediating and protective factors. Journal of Abnormal Child Psychology. [ePub ahead of print]. doi: 10.1007/s10802-017-0314-5.

[16] Kidziński, Ł., Yang, B., Hicks, J. L., Rajagopal, A., Delp, S. L., & Schwartz, M. H. (2020). Deep neural networks enable quantitative movement analysis using single-camera videos. Nature communications, 11(1), 1-10.

[17] I. S. Gracia, O. D. Suarez, G. B. Garcia, and T.-K. Kim, ‘‘Fast fight detection,’’ PLoS ONE, vol. 10, no. 4, Apr. 2015, Art. no. e0120448

[18] Zhou, P., Ding, Q., Luo, H., & Hou, X. (2018). Violence detection in surveillance video using low-level features. PLoS one, 13(10), e0203668.

[19] Carneiro, S. A., da Silva, G. P., Guimaraes, S. J. F., & Pedrini, H. (2019, October). Fight Detection in Video Sequences Based on Multi-Stream Convolutional Neural Networks. In 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI) (pp. 8-15). IEEE.

[20] Febin, I. P., Jayasree, K., & Joy, P. T. (2019). Violence detection in videos for an intelligent surveillance system using MoBSIFT and movement filtering algorithm. Pattern Analysis and Applications, 1-13.

[21] Datta, A., Shah, M., & Lobo, N. D. V. (2002, August). Person-on-person violence detection in video data. In Object recognition supported by user interaction for service robots (Vol. 1, pp. 433-438). IEEE.

[22] Nievas, E. B., Suarez, O. D., García, G. B., & Sukthankar, R. (2011, August). Violence detection in video using computer vision techniques. In International conference on Computer analysis of images and patterns (pp. 332-339). Springer, Berlin, Heidelberg.

[23] Zhou, P., Ding, Q., Luo, H., & Hou, X. (2017, June). Violent interaction detection in video based on deep learning. In Journal of physics: conference series (Vol. 844, No. 1, p. 012044). IOP Publishing.

[24] Pawar, K., & Attar, V. (2019). Deep learning approaches for video-based anomalous activity detection. World Wide Web, 22(2), 571-601.

[25] Ullah, F. U. M., Ullah, A., Muhammad, K., Haq, I. U., & Baik, S. W. (2019). Violence detection using spatiotemporal features with 3D convolutional neural network. Sensors, 19(11), 2472.

[26] Bauman, S. (2016). Do we need more measures of bullying?. Journal of Adolescent Health, 59(5), 487-488.

[27] Solberg M.E., Olweus D. Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire //Aggressive Behavior: Official Journal of the International Society for Research on Aggression. – 2003. – Т. 29. – №. 3. – С. 239-268.

[28] Shetgiri R. Bullying and victimization among children //Advances in pediatrics. – 2013. – Т. 60. – №. 1. – С. 33.

[29] Lloyd, K., Rosin, P. L., Marshall, D., & Moore, S. C. (2017). Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (GLCM)-based texture measures. Machine Vision and Applications, 28(3-4), 361-371.

[30] Bilinski, P., & Bremond, F. (2016, August). Human violence recognition and detection in surveillance videos. In 2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 30-36). IEEE.

[31] Yan, M., Meng, J., Zhou, C., Tu, Z., Tan, Y. P., & Yuan, J. (2020). Detecting spatiotemporal irregularities in videos via a 3D convolutional autoencoder. Journal of Visual Communication and Image Representation, 67, 102747.

[32] R. Olmos, S. Tabik, and F. Herrera, ‘‘Automatic handgun detection alarm in videos using deep learning,’’ Neurocomputing, vol. 275, pp. 66–72, Jan. 2018

[33] I. S. Gracia, O. D. Suarez, G. B. Garcia, and T.-K. Kim, ‘‘Fast fight detection,’’ PLoS ONE, vol. 10, no. 4, Apr. 2015, Art. no. e0120448.

[34] P. C. Ribeiro, R. Audigier, and Q. C. Pham, ‘‘RIMOC, a feature to discriminate unstructured motions: Application to violence detection for videosurveillance,’’ Comput. Vis. Image Understand., vol. 144, pp. 121–143, Mar. 2016

[35] V. E. M. Arceda, K. M. F. Fabián, P. C. L. Laura, J. J. R. Tito, and J. C. G. Cáceres, ‘‘Fast face detection in violent video scenes,’’ Electron. Notes Theor. Comput. Sci., vol. 329, pp. 5–26, Dec. 2016.

[36] J. Xie, W. Yan, C. Mu, T. Liu, P. Li, and S. Yan, ‘‘Recognizing violent activity without decoding video streams,’’ Optik, vol. 127, no. 2, pp. 795–801, Jan. 2016

[37] E. Y. Fu, H. Va Leong, G. Ngai, and S. Chan, ‘‘Automatic fight detection in surveillance videos,’’ in Proc. 14th Int. Conf. Adv. Mobile Comput. Multi Media, Nov. 2016, pp. 225–234.

[38] T. Senst, V. Eiselein, A. Kuhn, and T. Sikora, ‘‘Crowd violence detection using global motion-compensated Lagrangian features and scale-sensitive video-level representation,’’ IEEE Trans. Inf. Forensics Security, vol. 12, no. 12, pp. 2945–2956, Dec. 2017.

[39] S. Chaudhary, M. A. Khan, and C. Bhatnagar, ‘‘Multiple anomalous activity detection in videos,’’ Procedia Comput. Sci., vol. 125, pp. 336–345, Jan. 2018.

[40] E. Y. Fu, M. X. Huang, H. Va Leong, and G. Ngai, ‘‘Cross-species learning: A low-cost approach to learning human fight from animal fight,’’ in Proc. 26th ACM Int. Conf. Multimedia, Oct. 2018, pp. 320–327.

[41] M. Coletto, C. Lucchese, and S. Orlando, ‘‘Do violent people smile: Social media analysis of their profile pictures,’’ in Proc. Companion Proc. Web Conf., Apr. 2018, pp. 1465–1468.

[42] A. B. Mabrouk and E. Zagrouba, ‘‘Abnormal behavior recognition for intelligent video surveillance systems: A review,’’ Expert Syst. Appl., vol. 91, pp. 480–491, Jan. 2018.

[43] L. Auria and R. Moro, ‘‘Advantages and disadvantages of support vector machines,’’ in Credit Risk Assessment Revisited: Methodological Issues and Practical Implications, 2007, pp. 49–68

[44] T. Hassner, Y. Itcher, and O. Kliper-Gross, ‘‘Violent flows: Real-time detection of violent crowd behavior,’’ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 1–6.

[45] 3 - O. Deniz, I. Serrano, G. Bueno, and T.-K. Kim, ‘‘Fast violence detection in video,’’ in Proc. Int. Conf. Comput. Vis. Theory Appl. (VISAPP), vol. 2, Jan. 2014, pp. 478–485

[46] T. Zhang, Z. Yang, W. Jia, B. Yang, J. Yang, and X. He, ‘‘A new method for violence detection in surveillance scenes,’’ Multimedia Tools Appl., vol. 75, no. 12, pp. 7327–7349, 2016.

[47] X. Li, Y. Huo, Q. Jin, and J. Xu, ‘‘Detecting violence in video using subclasses,’’ in Proc. 24th ACM Int. Conf. Multimedia, Oct. 2016, pp. 586–590.

[48] P. Bilinski and F. Bremond, ‘‘Human violence recognition and detection in surveillance videos,’’ in Proc. 13th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Aug. 2016, pp. 30–36.

[49] Y. Gao, H. Liu, X. Sun, C. Wang, and Y. Liu, ‘‘Violence detection using oriented violent flows,’’ Image Vis. Comput., vol. 48, pp. 37–41, Apr./May 2016.

[50] C. Dhiman and D. K. Vishwakarma, ‘‘High dimensional abnormal human activity recognition using histogram oriented gradients and zernike moments,’’ in Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Dec. 2017, pp. 1–4.

[51] M. Al-Nawashi, O. M. Al-Hazaimeh, and M. Saraee, ‘‘A novel framework for intelligent surveillance system based on abnormal human activity detection in academic environments,’’ Neural Comput. Appl., vol. 28, no. 1, pp. 565–572, Dec. 2017.

[52] D. Song, C. Kim, and S.-K. Park, ‘‘A multi-temporal framework for highlevel activity analysis: Violent event detection in visual surveillance,’’ Inf. Sci., vol. 447, pp. 83–103, Jun. 2018.

[53] Q. Xia, P. Zhang, J. Wang, M. Tian, and C. Fei, ‘‘Real time violence detection based on deep spatio-temporal features,’’ in Proc. Chin. Conf. Biometric Recognit., 2018, pp. 157–165.

[54] B. M. Peixoto, S. Avila, Z. Dias, and A. Rocha, ‘‘Breaking down violence: A deep-learning strategy to model and classify violence in videos,’’ in Proc. 13th Int. Conf. Availability, Rel. Secur., Aug. 2018, Art. no. 50.

[55] A. A. Mishra and G. Srinivasa, ‘‘Automated detection of fighting styles using localized action features,’’ in Proc. 2nd Int. Conf. Inventive Syst. Control (ICISC), Jan. 2018, pp. 1385–1389.

[56] M.-Y. Chen and A. Hauptmann, ‘‘MoSIFT: Recognizing human actions in surveillance videos,’’ Figshare. Journal Contribution, 2018. doi: 10.1184/R1/6607523.v1

[57] T. Agrawal, A. Kumar, and S. K. Saraswat, ‘‘Comparative analysis of convolutional codes based on ML decoding,’’ in Proc. 2nd Int. Conf. Commun. Control Intell. Syst. (CCIS), Nov. 2016, pp. 41–45.

[58] C. Ding, S. Fan, M. Zhu, W. Feng, and B. Jia, ‘‘Violence detection in video by using 3D convolutional neural networks,’’ in Proc. Int. Symp. Visual Comput., 2014, pp. 551–558.

[59] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, ‘‘NetVLAD: CNN Architecture for weakly supervised place recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 5297–5307.

[60] G. Mu, H. Cao, and Q. Jin, ‘‘Violent scene detection using convolutional neural networks and deep audio features,’’ in Proc. Chin. Conf. Pattern Recognit., 2016, pp. 451–463.

[61] S. Sudhakaran and O. Lanz, ‘‘Learning to detect violent videos using convolutional long short-term memory,’’ in Proc. 14th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Aug./Sep. 2017, pp. 1–6.

[62] Z. Meng, J. Yuan, and Z. Li, ‘‘Trajectory-pooled deep convolutional networks for violence detection in videos,’’ in Proc. Int. Conf. Comput. Vis. Syst., 2017, pp. 437–447.

[63] I. Serrano, O. Deniz, J. L. Espinosa-Aranda, and G. Bueno, ‘‘Fight recognition in video using Hough forests and 2D convolutional neural network,’’ IEEE Trans. Image Process., vol. 27, no. 10, pp. 4787–4797, Oct. 2018.

[64] F. U. M. Ullah, A. Ullah, K. Muhammad, I. U. Haq, and S. W. Baik, ‘‘Violence detection using spatiotemporal features with 3D convolutional neural network,’’ Sensors, vol. 19, no. 11, p. 2472, May 2019.

[65] Green, R. D., & Guan, L. (2004). Quantifying and recognizing human movement patterns from monocular video images-part i: a new framework for modeling human motion. IEEE Transactions on Circuits and Systems for Video Technology, 14(2), 179-190.

[66] Hwang, J., & Tani, J. (2017). Seamless integration and coordination of cognitive skills in humanoid robots: a deep learning approach. IEEE Transactions on Cognitive and Developmental Systems, 10(2), 345-358.

[67] Marinoiu, E., Zanfir, M., Olaru, V., & Sminchisescu, C. (2018). 3d human sensing, action and emotion recognition in robot assisted therapy of children with autism. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2158-2167).

[68] El-Ghaish, H., Hussein, M. E., Shoukry, A., & Onai, R. (2018). Human action recognition based on integrating body pose, part shape, and motion. IEEE Access, 6, 49040-49055.

[69] Song, S., Lan, C., Xing, J., Zeng, W., & Liu, J. (2016). An end-to-end spatio-temporal attention model for human action recognition from skeleton data. arXiv preprint arXiv:1611.06067.

[70] Arici, T., Celebi, S., Aydin, A. S., & Temiz, T. T. (2014). Robust gesture recognition using feature pre-processing and weighted dynamic time warping. Multimedia Tools and Applications, 72(3), 3045-3062.

[71] Ji, Y., Cheng, H., Zheng, Y., & Li, H. (2015). Learning contrastive feature distribution model for interaction recognition. Journal of Visual Communication and Image Representation, 33, 340-349.

[72] Huynh-The, T., Le, B. V., Lee, S., & Yoon, Y. (2016). Interactive activity recognition using pose-based spatio–temporal relation features and four-level Pachinko Allocation Model. Information Sciences, 369, 317-333.

[73] Charalampous, K., Kostavelis, I., Boukas, E., Amanatiadis, A., Nalpantidis, L., Emmanouilidis, C., & Gasteratos, A. (2015). Autonomous robot path planning techniques using cellular automata. In Robots and lattice automata (pp. 175-196). Springer, Cham.

[74] H. S. Koppula, R. Gupta, and A. Saxena, “Learning human activities and object affordances from rgb-d videos,” International Journal of Robotics Research, vol. 32, no. 8, pp. 951–970, 2013.

[75] C. Granata, A. Ibanez, and P. Bidaud, “Human activity-understanding: A multilayer approach combining body movements and contextual descriptors analysis,” International Journal of Advanced Robotic Systems, vol. 12, no. 7, 2015.

[76] V. Dutta and T. Zielinska, “Predicting the intention of human activities for real-time human-robot interaction (hri),” in ICRA, 2016.

[77] [http://hacs.csail.mit.edu](http://hacs.csail.mit.edu/)

[78] Zhao, H., Torralba, A., Torresani, L., Yan, Z. Hacs: Human action clips and segments dataset for recognition and temporal localization //Proceedings of the IEEE International Conference on Computer Vision. – 2019. – С. 8668-8678.

[79] http://moments.csail.mit.edu

[80] Monfort M., Andonian A., Zhou B., Ramakrishnan K., Bargal S. A., Yan T., Brown L., Fan Q., Gutfreund D., Vondrick C., Oliva A. Moments in time dataset: one million videos for event understanding //IEEE transactions on pattern analysis and machine intelligence. – 2019. – Т. 42. – №. 2. – С. 502-508.

[81] https://research.google.com/ava/

[82] Публикация: Li, A., Thotakuri, M., Ross, D. A., Carreira, J., Vostrikov, A., Zisserman, A. The AVA-Kinetics Localized Human Actions Video Dataset //arXiv preprint arXiv:2005.00214. – 2020.

[83] https://oops.cs.columbia.edu/

[84] Epstein D., Chen B., Vondrick C. Oops! Predicting Unintentional Action in Video //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2020. – С. 919-929.

[85] https://20bn.com/datasets/something-something/v2

[86] Goyal R., Kahou S. E., Michalski V., Materzyńska J., Westphal S., Kim H., Haenel V., Fruend I., Yianilos P., Mueller-Freitag M., Hoppe F., Thurau C., Bax I., Memisevic R. The "Something Something" Video Database for Learning and Evaluating Visual Common Sense //ICCV. – 2017. – Т. 1. – №. 4. – С. 5.

[87] http://activity-net.org/

[88] Caba Heilbron, F., Escorcia, V., Ghanem, B., Carlos Niebles, J. Activitynet: A large-scale video benchmark for human activity understanding //Proceedings of the ieee conference on computer vision and pattern recognition. – 2015. – С. 961-970.

[89] http://visilab.etsii.uclm.es/personas/oscar/FightDetection/index.html

[90] http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/

[91] http://www.openu.ac.il/home/hassner/data/violentflows/

[92] Hassner, T.; Itcher, Y.; Kliper-Gross, O. Violent flows: Real-time detection of violent crowd behavior. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Providence, RI, USA, 16–21 June 2012; pp. 1–6.

[93] Nievas, E.B.; Suarez, O.D.; García, G.B.; Sukthankar, R. Violence detection in video using computer vision techniques. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, Seville, Spain, 29–31 August 2011; pp. 332–339. [Google Scholar]

[94] T. Hassner, Y. Itcher, and O. Kliper-Gross. Violent Flows: Real-Time Detection of Violent Crowd Behavior. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2012.

[95] Claire-Helene Demarty, Bogdan Ionescu, Yu-Gang Jiang, Vu Lam Quang, Markus Schedl, and Cedric Penet, “Benchmarking violent scenes detection in movies,” in Content-Based Multimedia Indexing (CBMI), International Workshop. IEEE, 2014, pp. 1–6

[96] K. Yun, J. Honorio, D. Chattopadhyay, T. L. Berg, and D. Samaras, ‘‘Two-person interaction detection using body-pose features and multiple instance learning,’’ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 28–35

[97] R. B. Fisher, ‘‘The PETS04 surveillance ground-truth data sets,’’ in Proc. 6th IEEE Int. Workshop Perform. Eval. Tracking Surveill., May 2004, pp. 1–5

[98] S. Gaglio, G. Lo Re, and M. Morana, ‘‘Human activity recognition process using 3-D posture data,’’ IEEE Trans. Human-Mach. Syst., vol. 45, no. 5, pp. 586–597, Oct. 2014

[99] (2014). [Online]. Available: http://www.dicgim.unipa.it/networks/ndslab/KARD/

[100] Y. Baveye, E. Dellandrea, C. Chamaret, and L. Chen, ‘‘LIRIS-ACCEDE: A video database for affective content analysis,’’ IEEE Trans. Affective Comput., vol. 6, no. 1, pp. 43–55, Jan. 2015.

[101] https://liris-accede.ec-lyon.fr/

[102] K. Soomro, A. R. Zamir, and M. Shah, ‘‘UCF101: A dataset of 101 human actions classes from videos in the wild,’’ 2012, arXiv:1212.0402. [Online]. Available: https://arxiv.org/abs/1212.0402

[103] K. Soomro, A. R. Zamir, and M. Shah, ‘‘UCF101: A dataset of 101 human actions classes from videos in the wild,’’ 2012, arXiv:1212.0402. [Online]. Available: https://arxiv.org/abs/1212.0402

[104] https://www.crcv.ucf.edu/data/UCF101.php