

УДК 004.421  
DOI 10.17513/snt.40239

## ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ МОДЕЛЕЙ НЕЙРОСЕТИ LRCN+LSTM В ЗАДАЧАХ РАСПОЗНАВАНИЯ НАСИЛИЯ НА ВИДЕО

Горяев В.М., Мацаков Б.В.

*ФГБОУ ВО «Калмыцкий государственный университет имени Б.Б. Городовикова»,  
Элиста, e-mail: goryaeff@mail.ru*

В данном исследовании целью было изучение эффективности моделей объединенных долгосрочной рекуррентной сверточной сети и сети долгой краткосрочной памяти в задаче распознавания насилия в видеофрагментах. Проводится оценка способности этих моделей обнаруживать и классифицировать насилие на видео. Для обучения был создан пользовательский набор данных из 150 помеченных видео открытого набора данных Violence Detection in Videos и 50 собственных видеофрагментов, размеченных на два класса (насилие и ненасилие). Для обучения использовались два типа сетей долгой краткосрочной памяти: однонаправленная и двунаправленная сети. Далее рассматривается эффективность объединенной архитектуры в распознавании насилия на видео, на базе оценки способности этих моделей обнаруживать и классифицировать насильственный контент в видео, а также анализ влияния различных параметров на их точность и скорость. Результаты исследования показывают, что модели сети долгой краткосрочной памяти совместно с рекуррентной нейросетью могут эффективно распознавать насилие на видео, особенно при использовании длинных и коротких временных масштабов. Для проекта используется модель сверточной сети дальнего действия, которая сочетает в себе преимущества сверточных сетей и рекуррентных нейронных сетей для извлечения объектов из видеок кадров и анализа временных связей между ними. Модель сверточной сети дальнего действия с двунаправленной долговременной памятью оказалась более эффективна при получении как пространственной, так и временной информации, а также прошлого и будущего контекста места происшествия, что приводит к улучшению обнаружения признаков насилия. Это исследование способствует разработке более точных и эффективных систем видеоанализа, которые могут быть полезны в различных областях применения, таких как наблюдение и безопасность. Потери при обучении и валидации с использованием нейросетей с архитектурой двунаправленной памяти стремятся к нулю, что подтверждает точность и эффективность данного алгоритма.

**Ключевые слова:** рекуррентная нейронная сеть, компьютерное зрение, распознавание действий человека, сети с долговременной и кратковременной памятью и дальнего действия, двунаправленная сеть, распознавание насилия

## AN INVESTIGATION OF THE EFFECTIVENESS OF LONG-RANGE AND SHORT-TERM MEMORY MODELS LRCN+LSTM IN VIDEO VIOLENCE DETECTION TASKS

Goryaev V.M., Matsakov B.V.

*Kalmyk State University named after B.B. Gorodovikov, Elista, e-mail: goryaeff@mail.ru*

This study examines In this study, the aim was to investigate the effectiveness of combined long-term recurrent convolutional networks and long short-term memory networks in the task of violence recognition in video fragments. An evaluation was conducted on the ability of these models to detect and classify violence in videos. A custom dataset was created for training, consisting of 150 labeled videos from the open dataset "Violence Detection in Videos" and 50 of our own video fragments annotated into two classes (violence and non-violence). Two types of long short-term memory networks were used for training: unidirectional and bidirectional networks. The effectiveness of the combined architecture in recognizing violence in videos is examined based on the assessment of these models' ability to detect and classify violent content, as well as the analysis of the impact of various parameters on their accuracy and speed. The results indicate that LSTM models combined with recurrent neural networks can effectively recognize violence in videos, particularly when using both long and short temporal scales. The project employs a long-range convolutional network model that combines the advantages of convolutional networks and recurrent neural networks for object extraction from video frames and analysis of temporal relationships between them. The long-range convolutional network model with bidirectional long-term memory proved to be more effective in obtaining both spatial and temporal information, as well as past and future context of the incident, which leads to improved detection of violence indicators. This research contributes to the development of more accurate and efficient video analysis systems that can be beneficial in various application areas, such as surveillance and security. The losses during training and validation using neural networks with bidirectional memory architecture approach zero, confirming the accuracy and effectiveness of this algorithm.

**Keywords:** recurrent neural networks, computer vision, long- and short-term memory networks, long-range networks, and bidirectional networks are used for violence recognition

### Введение

Распознавание человеческих действий (HAR) – процесс определения типа человеческой активности в видеопотоке – в последние годы получил значительное раз-

витие, в основном благодаря достижениям в области технологий глубокого обучения и методов анализа данных. На эту эволюцию также повлияло развитие моделей обнаружения объектов. Некоторые из хорошо

известных передовых технологий, которые способствовали этому прогрессу, включают рекуррентные сверточные сети (RNN), которые предназначены для обнаружения объектов. Долгосрочные рекуррентные сверточные сети (LRCN) объединяют сверточные нейронные сети (CNN) с рекуррентными нейронными сетями для анализа последовательных данных, в том числе и видео. В данном исследовании используется архитектура нейронной сети с длительной кратковременной памятью (LSTM), которая представляет собой тип RNN, специально предназначенный для решения проблемы исчезающего градиента, с которой обычно сталкиваются традиционные сети такого типа. Архитектура LRCN + LSTM эффективно сочетает в себе возможности извлечения признаков CNN и возможность LSTM изучать последовательности, что делает ее высокоэффективной для задач визуального распознавания и анализа временных данных.

Распознавание действий человека – определение типа действия человека на видеопотоке значительно эволюционировало за последние годы, благодаря улучшенным технологиям глубокого обучения и методов анализа данных [1]. Эволюция моделей обнаружения объектов оказала влияние и на развитие HAR [2]. В список таких передовых технологий включают рекуррентную нейронную сеть [3], быстрые регионы с использованием сверточных нейронных сетей (R-CNN) [4], а также популярное в настоящее время семейство технологий YOLO. Современные методы анализа видео в основном основаны на глубокой нейронной сети, в данной работе предлагается одна из моделей долговременной рекуррентной сверточной сети (RNN), сочетающая в себе средство извлечения визуальных признаков с глубокой иерархией с моделью, которая может научиться распознавать и синтезировать временную динамику для задач, связанных с последовательными данными (входными или выходными данными), визуальный, лингвистический или какой-либо другой [5]. В RNN градиент ошибки рассчитывается как произведение градиентов ошибки по каждому параметру, однако при этом он может становиться слишком большим, что приводит к нестабильности обучения [6]. Это связано с тем, что RNN имеет сложную структуру, которая включает в себя рекуррентные связи между ячейками. Сети LRCN, использующие LSTM, предназначены для решения этой проблемы путем эффективного отслеживания долгосрочных зависимостей во входных последовательностях и решения проблемы исчезающего градиента, характерной для традиционных се-

тей RNN. Этому способствует уникальная архитектура LSTM, которая включает в себя ячейки памяти и стробирующий механизм, позволяющий выборочно запоминать и забывать информацию в течение длительного периода времени. При этом необходимо учитывать, что ячейка памяти (рис. 1) здесь действует как промежуточный блок памяти, состоящий из более простых узлов, соединенных по определенной схеме [7].

Целью исследования является оценка двух моделей LSTM при распознавании сцен насилия в видеороликах, сравнение их эффективности и изучение влияния гиперпараметрического обучения на их эффективность, учитывая, что набор данных, используемый для обучения нейронной сети, имеет решающее значение для анализа, поскольку он предоставляет необходимые размеченные данные для оценки возможностей модели в распознавании сложных паттернов в видеопоследовательностях.

## Материалы и методы исследования

### *Набор данных*

Доступность маркированных наборов данных для обучения моделей распознавания насильственных действий существенно ограничена по сравнению с наборами данных для общих задач компьютерного зрения. Этот недостаток может быть обусловлен рядом факторов, включая опасения по поводу конфиденциальности пациентов и отсутствие общепринятых процедур обмена медицинскими данными. Чтобы решить эту проблему, исследователи предлагают несколько подходов, включая использование простых методов увеличения на базе масштабирования наборов данных и генерацию синтетических данных с помощью генеративно-противостоящих сеток GAN.

В данном проекте для обучения был создан пользовательский набор данных из содержащего 150 помеченных видео открытого набора данных Violence Detection in Videos [8] и 50 собственных видеофрагментов двух классов (насилие и ненасилие), разделенных на фрагменты по 10 с каждый. В процессе работы кода алгоритм выделяет видеофайлы на тренировочную и тестовую части (80/20) в две размеченные папки.

### *Долговременная кратковременная память LSTM*

LSTM решает проблемы взрыва градиента и исчезновения градиента в длинных последовательностях. Он также использует гейтинг для управления каждым блоком LSTM. Важно то, что LSTM имеет на одно

состояние  $C$  (ячейку памяти) больше, чем обычная RNN. Текущее состояние ячейки  $C_t$  может быть обновлено по следующей формуле:

$$\begin{cases} \widetilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c), & C_t = f_t \times C(t-1) + i_t \times \widetilde{C}_t, \\ f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) & i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) & o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \end{cases} \quad (1)$$

где  $C_{t-1}$  – предыдущее состояние ячейки,  $\sigma$  – функция активации,  $W$  – весовая матрица,  $b$  – смещение.

Таблица 1

Модель с UniLSTM

Слой	Тип	Параметры	Форма выхода слоя
Conv2D	TimeDistributed	128	(None, 28, 28, 32)
MaxPooling2D	TimeDistributed	–	(None, 14, 14, 32)
Flatten	TimeDistributed	–	(None, 6272)
LSTM	–	256	(None, 64)
Dense	–	2	(None, 2)

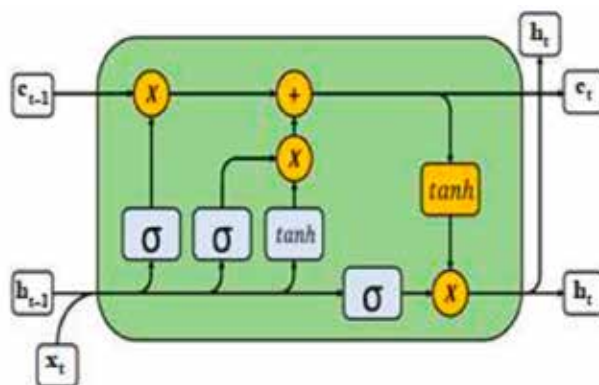


Рис. 1. Компоненты ячейки LSTM

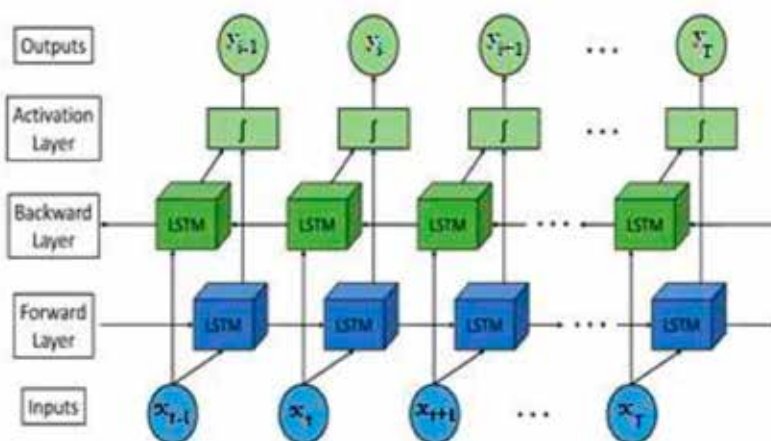


Рис. 2. Архитектура двунаправленной сети LSTM

Модель LRCN с однонаправленным LSTM (UniLSTM)

LRCN+LSTM – это модифицированная версия LRCN, которая включает в себя LSTM ячейки для улучшения способности

модели к запоминанию длинных последовательностей. Модель однонаправленной рекуррентной нейронной сети представляет собой архитектуру, которая использует одну рекуррентную нейронную сеть, работающую

щую в одном направлении. Для реализации модели однонаправленной рекуррентной нейронной сети используется API Keras. Каждый слой LSTM имеет 256 единиц, что соответствует размеру скрытого слоя и обрабатывает последовательные данные и генерирует 64-мерный результат. Этот результат затем классифицируется по одной из двух категорий с помощью плотного слоя (табл. 1).

Для обработки последовательных данных используется слой LSTM, а для сверточного и объединяющего слоев к каждому кадру видеопоследовательности используется оболочка TimeDistributed (рис. 1).

Таким образом, слой (вокруг которого он обернут) может принимать входные данные формы (no\_of\_frames, width, height, num\_of\_channels), если изначально входные данные слоя имели форму (width, height, num\_of\_channels).

*Модель LRCN с двунаправленным Bidirectional LSTM (BLSTM)*

BLSTM – это модифицированная версия LSTM, которая состоит из двух рекуррентных нейронных сетей, работающих в противоположных направлениях. Модель двунаправленной рекуррентной нейронной сети представляет собой архитектуру, которая сочетает в себе две рекуррентные сети, работающие в противоположных направлениях. Каждый двунаправленный слой LSTM имеет по 512 единиц, что соответствует размеру скрытого слоя. Каждый двунаправленный слой LSTM обрабатывает последовательные данные и генерирует 64-мерный результат, которые объединяются, чтобы получить 128-мерный вектор. Сам вектор затем классифицируется по одной из двух категорий с помощью плотного слоя (табл. 2). Используемый термин «глазок» основан на идее о том, что состояние ячейки «подматривает» за выходными данными ячейки, чтобы принимать решения о том, что делать с информацией. Это позволяет LSTM лучше фиксировать долгосрочные зависи-

мости в данных. Для вычисления выходных данных блока LRCN используется следующая формула:

$$h_t = A(x_t \times W_x h + h_{(t-1)} \times W_h h + b_h), \quad (2)$$

где  $A$  – функция активации,  $W$  – весовая матрица,  $b$  – смещение.

Скрытое состояние в момент времени  $t$  задается комбинацией  $H_t$ , выходной сигнал в любом заданном скрытом состоянии  $Y_t$ . Для обучения использовались два типа сетей долгой краткосрочной памяти. При реализации двунаправленной сети долгой краткосрочной памяти BLSTM поверх стандартного UniLSTM, был использован следующий подход: сначала идет прямой переход, где обработка входной последовательности идет стандартным способом, используя однонаправленный LSTM (формула (1)). Это дает представление о скрытом состоянии на каждом временном шаге. Далее обратный переход, обработка последовательности ввода пройдет в обратном порядке, начиная с последнего временного шага и возвращаясь к первому временному шагу, используя ту же архитектуру LSTM, что и при прямой технологии, но с обратной последовательностью ввода. Объединение скрытых состояний, полученных при прямом и обратном проходах, дает новое скрытое состояние, которое отражает как прошлый, так и будущий контекст.

На рис. 2 показана архитектура двунаправленной сети LSTM с (T) этапами [9]. Например, для момента (t) прямой слой использует данные момента (t-1) для генерации данных момента (t+1). Из уравнений (2) и (3) можно видеть, что в обратном слое данные момента (t+1) используются для генерации данных момента (t-1). Необходимо добавить, что оба процесса используют один и тот же слой активации независимо от направления потока информации. Далее можно описать работу двунаправленной рекуррентной нейронной сети.

**Таблица 2**

Модель с двунаправленной сетью LSTM

Слой	Тип	Параметры	Форма выхода слоя
Conv2D	TimeDistributed	128	(None, 28, 28, 32)
MaxPooling2D	TimeDistributed	–	(None, 14, 14, 32)
Flatten	TimeDistributed	–	(None, 6272)
Bidirectional LSTM	–	512	(None, 64)
Bidirectional LSTM	–	512	(None, 64)
Dense	–	2	(None, 2)

В сеть сначала вводится последовательность точек данных, каждая из которых представлена в виде вектора с одинаковой размерностью (рис. 1). При этом используется оболочка Bidirectional для создания двунаправленного LSTM-слоя с общими весами [10]. Аргумент merge\_mode определяет способ объединения прямых и обратных выходных данных в режиме конкатенации (рис. 2). Модель Bidirectional LSTM способна учитывать как предшествующий, так и последующий контекст в последовательных данных, данная задача оформлена в формуле (3). Это особенно полезно для таких задач, как языковое моделирование и видеоанализ. В процессе работы состояние ячейки «подсматривает» за выходными данными, что позволяет принимать более обоснованные решения о том, как обрабатывать поступающую информацию. Это позволяет сети лучше фиксировать долгосрочные зависимости в данных [11].

Для вычисления выходных данных блока LRCN используется следующая формула:

$$\begin{cases} H_t = \sigma(X_t \times W_{XH} + H_{t-1} + b_H), \\ H_t = \sigma(X_t \times W_{XH} + H_{t-1} \times W_{HH} + b_H), \end{cases} \quad (3)$$

где  $H_t$  – выходной сигнал на шаге  $t$ ,  $X_t$  – входной на шаге  $t$ ,  $W_{XH}$  – весовая матрица для входных сигналов,  $W_{HH}$  – весовая матрица для выходных,  $b_H$  – смещение,  $\sigma$  – активационная функция.

#### Результаты исследования и их обсуждение

В данном исследовании был проведен анализ производительности двух типов моделей долговременной кратковременной памяти: UniLSTM и BLSTM. Обе модели были обучены на наборе данных с видеороликами, содержащими насильственные и ненасильственные действия, с целью определения того, какая модель более эффективна при распознавании моделей и особенностей, связанных с насилием [12]. В процессе реализации проекта возникли проблемы при создании пользовательского набора данных, включая сложность переобучения и ограниченные вычислительные ресурсы, однако, несмотря на эти препятствия, в результате экспериментов с различными архитектурами, скоростью обучения, размерами пакетов и другими параметрами была определена оптимальная комбинация слоев и настроек. Обучение и тестирование проводились в течение 50 и 100 эпох соответственно. Производительность моделей оценивалась на основе нескольких показателей, включая потери, точность, потери при валидации и достоверность валидации.

Результаты показали, что BLSTM превосходит стандартную LSTM по точности и скорости обработки. Обучение как BLSTM, так и UniLSTM показало, что обе модели достигли высокой точности и низкого уровня потерь после 50 и 100 периодов. Однако BLSTM неизменно демонстрировала более высокую точность и меньшие потери по сравнению с однонаправленной моделью. После 50 эпох ее точность составила 0,85, а потери 0,3376, в то время как UniLSTM зафиксировала точность 0,845 и потерю 0,3207. Хотя разница в точности была минимальной, двунаправленная модель продемонстрировала более низкий уровень потерь. По прошествии 100 эпох точность обеих моделей улучшилась: Двунаправленная модель достигла 0,905, а однонаправленная – 0,889. BLSTM также продемонстрировала значительное снижение потерь, составив 0,3007 по сравнению с 0,3218 для однонаправленной модели. В итоге BLSTM превзошел UniLSTM как по точности, так и по потерям в обеих схемах.

Для оценки эффективности моделей были использованы показатели точности, позволяющие получить представление об их способности распознавать данные и прогнозировать результаты. Сравнительный анализ эффективности различных моделей обучения для выявления насилия и ненасилия в видеоклипах показал, что двунаправленная модель, обученная более 50 эпохам, продемонстрировала более высокую точность и полноту распознавания для обеих категорий по сравнению с однонаправленной.

Когда продолжительность обучения была увеличена до 100 периодов, двунаправленная модель продолжала демонстрировать улучшенную производительность. Для ненасильственных событий точность и полнота BLSTM сохранялись на уровне 0,89 и 0,77 соответственно, в то время как для UniLSTM точность снизилась до 0,67, а полнота – до 0,45. Напротив, для случаев насилия точность двунаправленной модели составила 0,90, а полнота – 0,91, что представляет собой значительное повышение по сравнению с точностью однонаправленной модели, равной 0,80 для случаев насильственного поведения (табл. 3). Графики, полученные в результате исследования, дают ценную информацию о производительности моделей LRCN+LSTM при распознавании насильственных и ненасильственных действий в видеоклипах (рис. 3). Далее, анализ матрицы ошибок показывает, что обе модели LSTM демонстрируют неплохие результаты в классификации насилия и ненасилия при увеличении количества эпох.

Таблица 3

## Отчет о классификации для сетей LRCN+LSTM

	Однонаправленное LSTM			Двунаправленное LSTM			
	точность (precision)	отзыв	f1-score	точность (precision)	отзыв	f1-score	Поддержка
epochs = 50							
Ненасилие	0,75	0,92	0,85	0,84	0,90	0,84	198
Насилие	0,94	0,71	0,81	0,89	0,81	0,83	202
Точность (accuracy)			0,83			0,84	400
avg*	0,76	0,95	0,85	0,86	0,86	0,85	400
wevg**	0,94	0,71	0,81	0,86	0,85	0,85	400
epochs = 100							
Ненасилие	0,96	0,67	0,79	0,90	0,89	0,89	198
Насилие	0,75	0,97	0,84	0,89	0,90	0,90	202
Точность (accuracy)			0,82			0,90	400
avg*	0,85	0,82	0,82	0,90	0,89	0,89	400
wevg**	0,85	0,82	0,82	0,90	0,90	0,89	400

Примечание. \* avg – среднее значение метрики для каждого класса; \*\* wevg – метрика, которая учитывает веса классов при расчете среднего значения метрики.

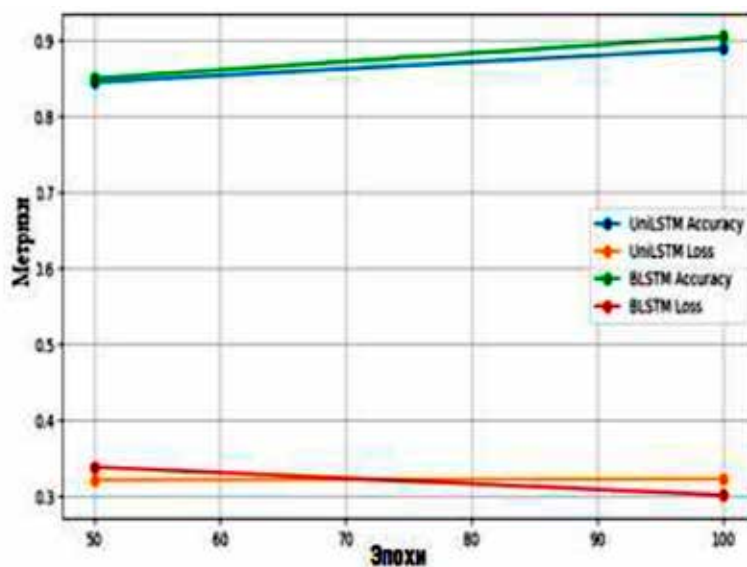


Рис. 3. Производительность моделей для разных параметров эпох

Таблица 4

## Матрица ошибок для 50 эпох

	Ненасилие	Насилие
UniLSTM		
Ненасилие	172	26
Насилие	51	149
BiLSTM		
Ненасилие	178	22
Насилие	36	164
	Ненасилие	Насилие

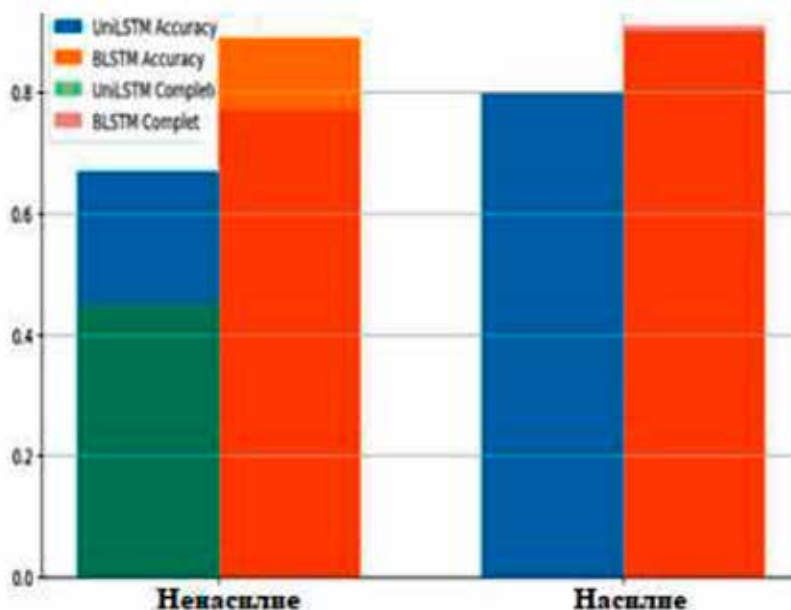


Рис. 4. Точность и полнота информации о насильственных и ненасильственных событиях

Таблица 5

Матрица ошибок для 100 эпох

	Ненасилие	Насилие
UniLSTM		
Ненасилие	188	10
Насилие	58	144
BLSTM		
Ненасилие	186	12
Насилие	40	162
	Ненасилие	Насилие

Однако BLSTM демонстрирует лучшие результаты в определенных сценариях (рис. 4). На 50 эпохах она показывает (табл. 4) лучшие результаты в распознавании насилия (82,0% против 74,3% для UniLSTM) и ненасилия (90,8% против 87,4% для UniLSTM). Кроме того, на 100 эпохах BLSTM также демонстрирует лучшие результаты в распознавании насилия (88,3%) и ненасилия (96,4%). В целом двунаправленная модель демонстрирует преимущества в распознавании насилия и ненасилия, особенно на 100 эпохах (табл. 5).

Это может быть связано с тем преимуществом, что может использовать информацию из обоих направлений последовательности, что позволяет ему лучше понимать контекст и распознавать насилие. В исследовании сравниваются две модели для LRCN + LSTM: UniLSTM и BLSTM, которые используются для распознавания на-

силия и ненасилия в видеороликах. BLSTM демонстрирует лучшие результаты по точности и скорости обработки по сравнению с UniLSTM. После 50 эпох двунаправленная модель показывает точность 0,85 и потери 0,3376, в то время как однонаправленная модель показывает точность 0,845 и потери 0,3207. После 100 эпох сеть BLSTM увеличила точность до 0,905 и, соответственно, уменьшила потери до 0,3007. На 50 эпохах двунаправленная модель показывает лучшие результаты в распознавании насилия (82,0% против 74,3% для UniLSTM) и ненасилия (90,8% против 87,4% для UniLSTM). На 100 эпохах BLSTM также демонстрирует лучшие результаты в распознавании насилия (88,3%) и ненасилия (96,4%).

Обе модели демонстрируют достаточную способность распознавать ненасильственные действия, но у BLSTM немного выше точность и меньше ошибок классификации. Итоги работы подтверждают, что двунаправленная модель является более эффективной моделью для распознавания насилия и ненасилия в видеофрагментах, чем однонаправленная, это важно для разработки систем распознавания признаков насилия, которые могут быть использованы в различных приложениях, таких как мониторинг видео и анализ поведения.

### Заключение

Несмотря на растущую доступность крупномасштабных наборов данных для общих задач компьютерного зрения, аннотированные наборы данных для моделей рас-

познавания насилия по-прежнему мало представлены. Согласно отчету *International Journal of Computer Vision*, только 10% общедоступных наборов данных для задач компьютерного зрения специально ориентированы на распознавание насилия и многие из этих наборов данных невелики и необъективны.

Проведенное исследование показало, что двунаправленная модель долговременной кратковременной памяти (BLSTM) является более эффективной моделью для распознавания насилия и ненасилия в видеофрагментах, чем однонаправленная модель (UniLSTM). Результаты экспериментов показали, что BLSTM демонстрирует лучшие результаты в распознавании насилия (88%) и ненасилия (96%). Эти результаты имеют значения для разработки систем распознавания признаков насилия, которые могут быть использованы в различных приложениях, таких как мониторинг видео и анализ поведения. Кроме того, результаты исследования могут быть использованы для улучшения существующих систем распознавания насилия и ненасилия. Несмотря на это, результаты данного исследования демонстрируют потенциал моделей LSTM для распознавания насилия в видеороликах, надо учесть, что оценка модели на основе небольшого набора данных может привести к переобучению, подбору и неадекватной оценке модели, что, в свою очередь, может вызвать неадекватную оценку точности модели и ее способности справиться с новыми данными.

#### Список литературы

1. Tan H.H., Lim K.H. Vanishing Gradient Mitigation with Deep Learning Neural Network Optimization // 7th International Conference on Smart Computing & Communications (ICSCC). Sarawak, Malaysia. 2019. P. 1–4. DOI: 10.1109/ICSCC.2019.8843652.
2. Zitnick S., Lawrence C., Dollar P. Edge Boxes. Locating Object Proposals from Edges // *Computer Vision-ECCV 2014*. Springer International Publishing. 2014. P. 391–405. DOI: 10.1007/978-3-319-10602-1\_26.
3. Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017. Vol. 39, Is. 6. P. 1137–1149. DOI: 10.48550/arXiv.1506.01497.
4. Goryaev V.M., Basangova E.O. et al. Forecasting steppe fires using remote sensing data of time series // *IOP Conference Series: MSE*. 2021. Vol. 1047, Is. 1. P. 12092–12098. DOI: 10.1088/1757-899X/1047/1/012092.
5. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas NV USA. 2016. P. 770–778. DOI: 10.1109/CVPR.2016.90.
6. Hochreiter S., Schmidhuber J. Long short-term memory // *Neural Computation*. 1997. Vol. 9, Is. 8. P. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
7. Goryaev V.M., Uchurova E.O., Basangova E.O., Bembitov D.B., Miloshenko A.P. Analysis of digital filters for preprocessing biomedical signals from ECG apparatus // *AIP Conference (MIST: Aerospace-IV)*. AIP publishing. 2023. Vol. 2700 (1). P. 050037–050043. DOI: 10.1063/5.0125057.
8. Graves A., Schmidhuber J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures // *Neural networks*. 2005. Vol. 18, Is. 5–6. P. 602–610. DOI: 10.1016/j.neunet.2005.06.042.
9. Guan Y. Abnormal behavior recognition using 3D-CNN combined with LSTM // *Multimedia Tools and Applications*. 2021. Vol. 80, Is. 8. P. 18787–18801. DOI: 10.1007/s11042-021-10667-9.
10. Ling Z.-H., Kang S.-Y., Zen H. Deep learning for acoustic modeling in parametric speech generation: A systematic review of existing techniques and future trends // *IEEE Signal Processing Magazine*. 2015. Vol. 32, Is. 3. P. 35–52. DOI: 10.1109/MSP.2014.2359987.
11. Kim Ji., Kim Ja., Le T., Kim H. Long Short-Term Memory Recurrent Neural Network Classifier for Intrusion Detection // *International Conference on Platform Technology and (PlatCon)*, Jeju Korea. 2016. Is. 2. P. 1–5. DOI: 10.1109/PlatCon.2016.7456805.
12. Мацаков Б.В., Горяев В.М. Распознавание признаков насилия с помощью сетей долгой краткосрочной памяти LSTM // Свидетельство о государственной регистрации программы для ЭВМ 2024690113 Российская Федерация. № 2024688048: заявл. 15.11.2024; опубли. 12.12.2024 / заявитель и правообладатель В.М. Горяев. 1 с.