

*Тронов Кирилл Александрович, студент-магистр, Калужский филиал ФГБОУ*

*ВО «Московский государственный технический университет имени Н.Э.*

*Баумана (национальный исследовательский университет)»*

*Белов Юрий Сергеевич, к.ф. -м.н., доцент, Калужский филиал ФГБОУ ВО*

*«Московский государственный технический университет имени Н.Э. Баумана*

*(национальный исследовательский университет)»*

## **НЕЙРОННЫЕ СЕТИ MULTI-CONV LSTM ДЛЯ РАСПОЗНАВАНИЯ АКТИВНОСТИ ЧЕЛОВЕКА НА ОСНОВЕ ДАТЧИКОВ СМАРТФОНА**

**Аннотация:** В последние годы распознавание человеческой деятельности привлекает большое внимание благодаря своему широкому применению, которое отчасти обеспечивается за счет широкого распространения смартфонов, оснащенных стандартным набором датчиков. Нейронная сеть Multi-convLSTM позволяет эффективно решать задачу распознавания благодаря сочетанию сверточных сетей, предоставляющих возможность автоматического и адаптивного извлечения релевантных и надежных характеристик и LSTM, позволяющих извлекать внутренние временные особенности из зависимости сигналов временных рядов.

**Ключевые слова:** распознавание активности человека, lstm, рекуррентные нейронные сети, рнс.

**Annotation:** In recent years, recognition of human activity has received a lot of attention due to its widespread use, which is partly provided by widespread adoption of smartphones equipped with a standard set of sensors. The Multi-convLSTM neural network allows you to effectively solve the recognition problem thanks to a combination of convolutional networks that provide the ability to automatically and adaptively extract relevant and reliable characteristics and LSTMs

that allow you to extract internal temporal features from the dependence of time series signals.

**Keywords:** recognition of human activity, lstm, recurrent neural networks, rnn.

**Введение.** В современных реалиях распознавание активности человека является довольно востребованной задачей. Наиболее подходящим устройством для данной задачи является смартфон за счет широкого распространения и оснащенности стандартным набором датчиков, позволяющим провести позиционирование, однако сам алгоритм распознавания сложнее, чем при наличии фиксированных датчиков запястья, талии и ног. Нейронная сеть Multi-convLSTM позволяет эффективно решать данную задачу благодаря сочетанию сверточных и рекуррентных нейронных сетей, а задача обнаружения признаков решается глубоким обучением [1]. В данной статье предлагается глубокая сетевая структура на основе MconvLSTM для распознавания человеческой активности в реальном времени, с особым подходом к расширению данных и методом построения сети.

**Предложенный метод.** Рекуррентная нейронная сеть (РНС) осуществляет самообратную связь, задерживая активацию значений ячеек, что позволяет изучать особенности временных рядов данных. Однако признаки, находящиеся на большом расстоянии друг от друга, не могут быть скоррелированы с помощью RNN в последовательности на большом расстоянии. LSTM (Long Short Term Memory сети) эффективно решает эту проблему. LSTM преобразует нейроны в четыре сетевых блока, которые управляются фильтрами состояния: запись, чтение и сброс. Сверточные нейронные сети (CNN) обычно используются для извлечения пространственных особенностей из входных сенсорных данных. Одномерное ядро свертки используется для фильтрации выбросов и извлечения признаков.

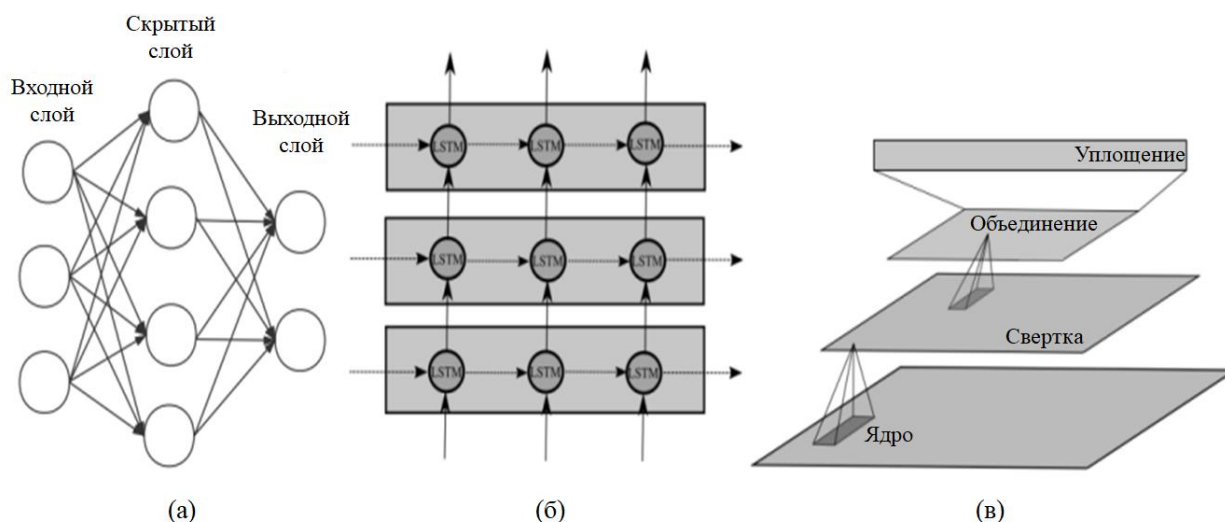


Рис.1 Модели, использованные в данной работе. Слева направо: (а) Полностью подключенная сеть с прямой связью (б) Скрытые слои сети LSTM, содержащие ячейки LSTM (в) сверточные сети, содержащие слои сверток, объединения и уплощения

Проблему распознавания человеческой деятельности можно рассматривать как проблема распознавания образов, а predetermined набор действий  $A$  выражается как:

$$A = \{a_i\}_{i=1}^m,$$

где  $m$  - тип набора активности. В то же время, информация о времени работы датчика  $S$  выражается как:

$$s_n^k = \{d_1^k, d_2^k, d_3^k, \dots, d_n^k\}, s_n^k \in S.$$

Данные собираются трехкоординатным датчиком, где  $k$  - размер датчика, а  $d \in \mathbb{N}$  представляет собой  $k$ -мерные данные, считанные в момент времени  $t$ . Постройте модель  $F$  и информацию о временной последовательности  $S$ , чтобы определить типы человеческой деятельности, а именно:

$$\hat{A} = \{a_j\}_{j=1}^m = F(S), \hat{A} \in A(4)$$

На рис. 2 показана система распознавания человеческой активности на основе модели MconvLSTM. Собранные сенсорные данные подвергаются предварительной обработке. Затем соответствующие модули ConvLSTM строятся в соответствии с количеством сенсорных данных, где каждый модуль содержит три сверточных слоя и три блока LSTM, вывод каждого модуля

ConvLSTM параллельно помещается в полное соединение, и, наконец, определяется категория через Softmax.

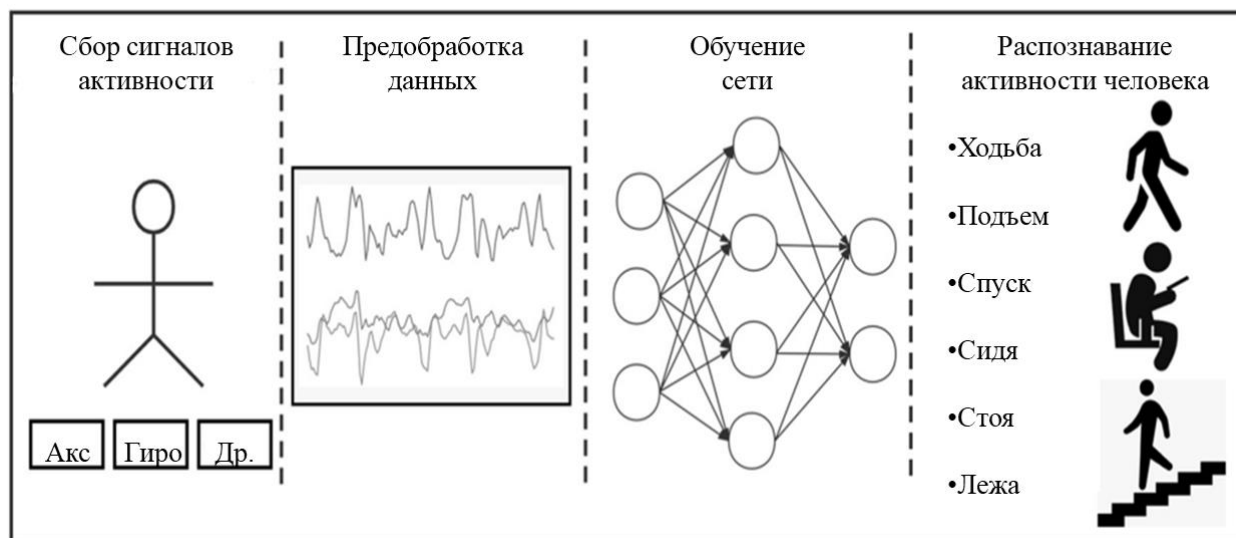


Рис.2 Иллюстрация распознавания активности на основе сенсоров с использованием подходов MconvLSTM

Инерциальный измерительный блок (IMU) в смартфоне включает акселерометр и гироскоп. Комбинированное измерение нескольких датчиков позволяет получить более точный эффект классификации по сравнению с одним датчиком [2]. Датчик разделен на три оси (x, y, z) для сбора данных. Для входных данных  $S$  стандартизируется один сигнал:

$$S_k^{ij} = \sqrt{\sum_{t=i}^j S_k^t * S_k^t}$$

где  $k$  - тип датчика, а  $S$  - входные данные датчика. Стандартизированные данные помещаются во вновь построенную последовательность. По мере увеличения размерности все расширенные последовательности будут помещены в новую последовательность [3]. Сенсорный модуль MconvLSTM включает три типа структур, а именно: входной слой, несколько слоев свертки и слои LSTM [4]. Входной слой вводит данные датчика в модель в скользящем окне. Сверточный слой используется для извлечения пространственных

признаков. Активация или выход сверточного соединения вычисляется следующим образом:

$$c_i^{l,j} = \sigma\left(\sum_{m=1}^M \psi_m^{l,j} * x_{i+m-1}^{l,j} + b_j^i\right)$$

где  $l$  - количество слоев,  $\sigma$  - функция активации,  $\psi_m^j$  -  $m$ -е ядро свертки с признаком  $j$ ,  $x$  - входная единица длины  $N$ , выведенная из предыдущего слоя,  $b_j^i$  смещение  $j$ -го признака. Для повышения скорости сходимости сети входной вектор  $b_i^{l,j}$  помещается в слой стандартизации после слоя свертки [5]:

$$n_i^{l,j} = \frac{c_i^l - \mu_c}{\sqrt{\varepsilon^2 + \varepsilon}}$$

где  $\mu_c$  и  $\varepsilon$  соответственно представляют среднее и стандартное отклонение входного вектора, и затем помещаются в слой объединения для понижающей выборки после нормализации. Слой объединения извлекает особенности из  $m_i^{l,j}$ , и затем используется maxpooling для вывода максимального значения входного вектора следующим образом:

$$p_i^{l,j} = \max_{r \in R} (n_{i * T + r}^{l,j})$$

где  $R$  - размер объединения, а  $T$  - размер скользящего шага. Несколько сверточных слоев, стандартных слоев и слоев объединения образуют структуру нейронной решетки. Эта структура сети извлекает признаки через слои, и более сложные признаки могут быть извлечены по мере углубления слоя сети. После построения многослойной сверточной сети она будет введена в блок LSTM, который обновляет статус временных характеристик. Наконец, вывод последовательности признаков одного датчика объединяется. В качестве входных данных для слоя LSTM:

$$h_t^{l,j} = \sigma(\omega_{t,h}^{l,j} x_t^{l,j} + h_{t-1}^{l,j} \omega_{t-1,h}^{l,j} + b_t^l)$$

где  $b_t^l$  представляет собой смещение блока,  $x_t^l$  - входной вектор для блока  $t$ , а  $\omega$  - весовой коэффициент каждого блока. Способ построения сети с одной последовательностью легко игнорирует неоднородность различных сенсорных

данных. Модуль сети convLSTM создается на основе количества типов датчиков, что позволяет увеличить их количество. После вывода модуля convLSTM, он параллельно строится на полностью подключённый слой:

$$h_t^l = \sigma(h_{t1}^{l,j} : h_{t2}^{l,j} : \dots : h_{tm}^{l,j})$$

Наконец, классификатор Softmax объединяет и идентифицирует категории деятельности и выводит результаты классификации. Классификация следующая:

$$P(\eta | h) = \arg \max \left( \frac{\exp(h^{L-1}w^L + b^L)}{\sum_{n=1}^N \exp(h^{L-1}w^L)} \right)$$

где  $\eta$  - конечная выходная категория,  $N$  - общее количество категорий активности,  $L$  - количество слоев. Для того чтобы ускорить обучение сети и снизить стоимость вычислений, используется оптимизатор Адама для обновления весов:

$$m_i = \beta_1 * m_{i-1} + (1 - \beta_1) dx$$

$$v_i = \beta_2 * v_{i-1} + (1 - \beta_2) dx^2$$

$$W_i = W_{i-1} + \frac{-\alpha m_i}{\sqrt{v_i}}$$

где  $\beta$  - скорость экспоненциального затухания,  $\alpha$  - скорость обучения, а  $W$  - матрица весов.

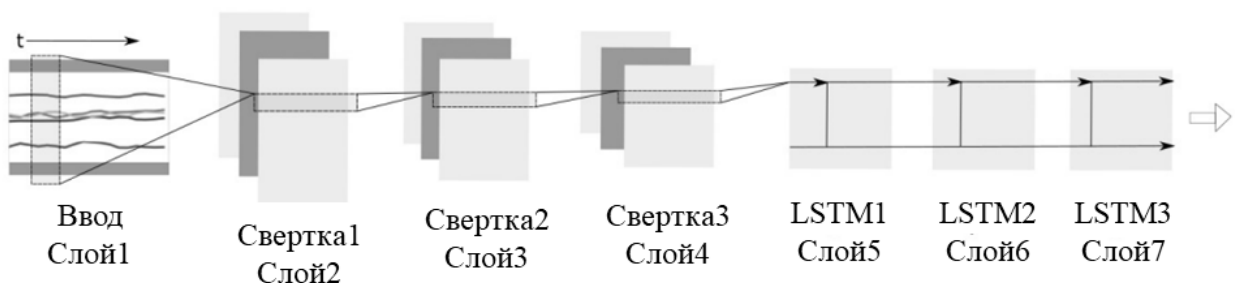


Рис.3 Структура сети MconvLSTM

**Заключение.** В данной статье был предложен метод расширения данных и построения сети MconvLSTM на основе входных данных от нескольких датчиков, с использованием акселерометра и гироскопа смартфона для эффективного распознавания человеческой активности. Данную эффективность обеспечивает сочетание сверточных сетей, предоставляющих возможность автоматического и адаптивного извлечения релевантных и надежных характеристик и LSTM, позволяющих извлекать внутренние временные особенности из зависимости сигналов временных рядов.

#### **Библиографический список:**

1. Chen K , Zhang D , Yao L , et al. Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges and Opportunities[J]. 2020.
2. Ростовцев В.С. Искусственные нейронные сети. Учебник – Санкт-Петербург: Лань, 2021. – 216с.
3. Wang J , Chen Y , Hao S , et al. Deep Learning for Sensor-based Activity Recognition: A Survey[J]. PatternRecognitionLetters, 2018.
4. Liu J, Shahroudy A , Xu D , et al. Spatio-Temporal LSTM with Trust Gates for 3D Human Action Recognition[J]. 2016.
5. Николенко С.И., Кадури А. А. Глубокое обучение. Погружение в мир нейронных сетей – Санкт-Петербург: Питер, 2020. – 480с.