

# Нейросетевой алгоритм для классификации наборов ключевых точек тела человека

И.А. Килбас  
Самарский национальный  
исследовательский университет им.  
академика С.П. Королева  
Самара, Россия  
igor.kilbas.ai@gmail.com

Д.Н. Грибанов  
Самарский национальный  
исследовательский университет им.  
академика С.П. Королева  
Самара, Россия  
cool.danik01@yandex.ru

Р.А. Парингер  
Самарский национальный  
исследовательский университет им.  
академика С.П. Королева  
Институт систем обработки  
изображений - филиал ФНИЦ  
«Кристаллография и фотоника»  
РАН  
Самара, Россия  
rusparinger@gmail.com

**Аннотация**—В данной работе представлен нейросетевой алгоритм классификации наборов ключевых точек тела человека. Особенностью данного алгоритма является подход к предобработке данных, позволяющий увеличить обобщающую способность нейронной сети, производящей классификацию. Результаты экспериментов показывают, что предложенный алгоритм обладает высокой точностью и может быть использован в качестве компонента в системах распознавания позы человека.

**Ключевые слова**— классификация позы, ключевые точки, нейронные сети, распознавание позы.

## 1. ВВЕДЕНИЕ

Задача понимания человека машинами представляет фундаментальное значение. Благодаря развитию нейросетевых алгоритмов был достигнут существенный прогресс в данном направлении. Особый интерес представляет распознавание позы человека на изображении, поскольку это является неотъемлемой частью понимания состояния человека машинами. Было предложено множество подходов [1-3], большое количество которых применяет нейронные сети.

Задача распознавания позы человека сводится к обнаружению (регрессии) ключевых точек на теле человека и их последующей группировке в отдельные наборы (скелеты). Далее в зависимости от конкретного приложения извлекаются разные типы информации из полученных скелетов: положение человека в кадре, угол наклона конкретной конечности, класс позы. В данной работе рассматривается последняя задача. Представлен нейросетевой алгоритм классификации наборов ключевых точек. Алгоритм разработан таким образом, чтобы быть устойчивым к различным форматам ключевых точек, а также к различного рода трансформациям: сдвиг точек, масштабирование точек.

## 2. НАБОР ОБУЧАЮЩИХ ДАННЫХ

Для обучения алгоритма был собран набор данных, представляющий собой множество наборов ключевых точек тела человека, размеченные 36 классами («руки на груди», «присед» и т.д.). Пример позы показан на рисунке 1. Классы представляют собой различные упражнения, выполняемые человеком. Наиболее сложными для распознавания классами являются упражнения, включающие касание колен или носков

руками. Далее данные классы будут называться классы позы типа D.

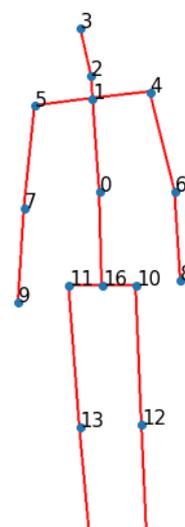


Рис. 1. Пример набора ключевых точек

## 3. ОПИСАНИЕ АЛГОРИТМА

Алгоритм состоит из двух модулей: модуль предобработки и нейронная сеть. Модуль предобработки отвечает за извлечение признаков из входного набора ключевых точек. Нейронная сеть же производит классификацию вектора извлеченных признаков. Далее будет представлено описание каждого из компонентов.

### А. Предобработка точек

Для обеспечения высокого уровня обобщающей способности нейронной сети входные признаки должны содержать только релевантную информацию для исходной задачи. В частности, они должны удовлетворять следующим требованиям:

- 1) Быть инвариантным к масштабу координат ключевых точек.
- 2) Быть инвариантным к сдвигу набора точек.

Для удовлетворения обоих требований на основе набора точек тела человека будет строиться набор нормированных векторов между заданными парами

точек. В качестве таких пар, например, можно взять точки, лежащие на конечностях человека (руки, ноги). Требование 1 выполняется, потому что вектора нормированы (имеют норму, равную единице). Требование 2 выполняется, потому что вектора представляют собой разности значений координат точек, а данная разность не зависит от абсолютного положения точек, делая вектора инвариантными к сдвигу. Помимо выполнения выше озвученных требований, данный набор векторов также инвариантен к пропорциям тела, поскольку не содержит в себе информации о длинах конечностей.

Несмотря на то, что описанный набор векторов удовлетворяет всем требованиям, он не способен запечатлеть наклон торса вперёд, поскольку направления векторов на конечностях не изменяются, меняются лишь их длины до нормировки. Для исправления этой проблемы дополнительно просчитывается относительное расстояние между руками и коленями, а также между руками и стопами. Данное расстояние относительно, поскольку оно будет нормироваться расстоянием между коленом и стопой, дабы не нарушать требование 2.

#### Б. Нейронная сеть

В качестве нейронной сети используется трёхслойная полносвязная нейронная сеть. Данная нейронная сеть вместе с модулем предобработки данных изображена на рисунке 2.

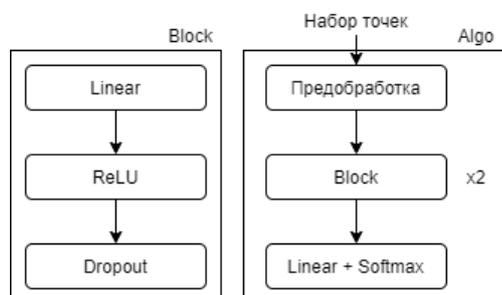


Рис. 2. Схема предложенного алгоритма

Нейронная сеть состоит из двух блоков, каждый блок состоит из слоёв полносвязного слоя (Linear), функции активации и регуляризирующего слоя Dropout. После каждого полносвязного слоя в качестве функции активации используется функция ReLU [4], уже ставшая стандартом во многих приложениях глубокого обучения. Для улучшения обобщающей способности нейронной сети используется Dropout [5], что случайным образом «выключает» нейроны во время обучения и тем самым регуляризирует нейронную сеть, предотвращая переобучение.

После блоков используется еще один полносвязный слой, производящий классификацию, и функция Softmax, что нормирует выходные значения данного слоя, конвертируя их в значения вероятностей принадлежности тому или иному классу.

## 4. РЕЗУЛЬТАТЫ

Для иллюстрации важности разработанной предобработки наборов ключевых точек было проведено сравнение разных версий алгоритма: исходный алгоритм – версия А, в модуле предобработки не вычисляются относительные расстояния – версия В, вместо модуля предобработки используется нормализация средним и дисперсией – версия С. Также были дополнительно рассчитаны значения точности по наиболее сложным классам поз (позы типа D). Значения точности разных версий алгоритма показаны в таблице 1.

Таблица 1. ЗНАЧЕНИЯ ТОЧНОСТИ КЛАССИФИКАЦИИ ПОЗ РАЗНЫМИ ВЕРСИЯМИ АЛГОРИТМА

Версия алгоритма	Средняя точность по всем классам поз	Средняя точность по классам поз типа D
А	89,8%	78,0%
В	87,2%	63,0%
С	79,9%	51,2%

Применение нейронной сети без модуля предобработки приводит к наименьшей точности, в частности на классах поз типа D. Использование предложенной предобработки, наоборот, позволяет получить наилучшую точность, в том числе более высокую точность на сложных классах поз, что подтверждает её важность.

## 5. ЗАКЛЮЧЕНИЕ

В данной работе был разработан и исследован нейросетевой алгоритм классификации наборов ключевых точек тела человека. Алгоритм представляет собой многослойную полносвязную нейронную сеть и модуль предобработки входных данных, что делает нейронную сеть инвариантной к ряду разного рода информации во входных данных, тем самым увеличивая её обобщающую способность. Исходя из результатов можно заключить, что предложенный алгоритм классификации наборов ключевых точек может быть использован в качестве компонента в системах распознавания позы человека.

## ЛИТЕРАТУРА

- [1] Cao, Z. Realtime multi-person 2d pose estimation using part affinity fields / Z. Cao, T. Simon, S. Wei, Y. Sheikh // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2017. – P. 7291-7299.
- [2] Artacho, B. OmniPose: A Multi-Scale Framework for Multi-Person Pose Estimation / B. Artacho, A. Savakis // ArXiv, 2021.
- [3] Jin, S. Whole-Body Human Pose Estimation in the Wild / S. Jin // Proceedings of the European Conference on Computer Vision (ECCV), 2020.
- [4] He, K. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification / K. He, X. Zhang, S. Ren, J. Sun // Proceedings of the IEEE International Conference on Computer Vision. – 2015. – P. 1026-1034.
- [5] Srivastava, N. Dropout: a simple way to prevent neural networks from overfitting / N. Srivastava // The journal of machine learning research. – 2014. – Vol. 15(1). – P. 1929-1958.