

Ю.А. Кева¹, Д.Е. Конева¹, Ю.А.Егоров^{1,2}

¹ Тюменский государственный университет, г. Тюмень

² Научно-технический университет «Сириус», г.Сочи

УДК 004.85

ИССЛЕДОВАНИЕ МЕТОДОВ КЛАССИФИКАЦИИ ПОЗ ЧЕЛОВЕКА С ПОМОЩЬЮ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Аннотация. В статье рассмотрено использование методов машинного и глубокого обучения для классификации бегущих и идущих людей на отдельных кадрах.

Ключевые слова: классификация поз человека, бегущие люди, идущие люди, машинное обучение, глубокое обучение, нейронные сети.

Введение

В данной работе рассмотрены методы классификации поз человека с помощью искусственных нейронных сетей.

В настоящее время в различных крупных аэропортах, местах большого скопления людей, даже в школах во время проведения экзаменов используются камеры для видеонаблюдения, записи происходящего. Подобные системы важны для обеспечения безопасности в людных местах, поэтому они продолжают совершенствоваться и постепенно выходят на новый уровень: вместо обычной записи видео начинают применяться методы распознавания людей, что впоследствии упрощает работу со снятым материалом.

Подобный подход можно также использовать в образовательном процессе, чтобы, например, отслеживать, чем ученики заняты во время урока, кто внимательно слушает преподавателя, а кто отвлекся и смотрит в окно или в телефон. Также распознавание поз человека может быть

применено в каких-либо спортивных мероприятиях и тренировках, где есть необходимость отслеживать действия людей в целях соблюдения ими правил или для составления индивидуального графика занятий.

Таким образом, возникает необходимость создать такую обученную модель, которая будет способна классифицировать позы человека, причем за короткий промежуток времени на видеозаписи.

Целью данного проекта было проверить, можно ли использовать нейронные сети для классификации бегущих и идущих людей на отдельных кадрах.

Обзор

При подготовке к обучению нейронной сети возникает необходимость создания обучающей выборки. Для задачи классификации поз человека качество выборки является важным фактором, влияющим на точность классификации у будущей модели. Готовую выборку подобрать не так просто. В интернете можно найти множество различных наборов данных для этой задачи в общем, но из них достаточно сложно выделить фотографии для классификации только бегущих и идущих людей.

Классификация поз при беге и ходьбе осложняется тем, что очень многие факторы влияют на обучение модели. Во-первых, речь идет о ракурсах съемки. Например, человеческому глазу несложно отличить бегущих и идущих людей в профиль, но задача осложняется, если смотреть на приближающегося к вам человека или удаляющегося от вас, где гораздо труднее рассмотреть положение рук и ног, характерных для бега или ходьбы. Во-вторых, влияние могут оказывать различные аксессуары, такие как рюкзаки и сумки. В-третьих, это одежда, которая отличается в зависимости от времени года. Например, длинные плащи, пуховики могут

скрывать движения и усложнять распознавание. В приведенной работе подробно рассмотрены факторы, влияющие на распознавание движений [3].

В качестве наборов данных для формирования обучающей выборки были рассмотрены две базы данных. Проблемой таких готовых наборов является то, что они зачастую содержат данные, которые потом требуют дополнительной обработки. К тому же эти два набора направлены на распознавание поз человека в целом, поэтому перед использованием необходимо выбрать из них данные для более узкой задачи классификации бегущих и идущих людей, из-за чего конечная обучающая выборка получается меньше.

Первым был рассмотрен набор из шести категорий видеозаписей [5]. В данном наборе содержится по 100 видео для каждой категории, видеозаписи примерно по 20 секунд. Для каждой записи снимался только один человек с определенного ракурса, из-за чего кадры из видео удобно использовать в качестве обучающей выборки. Однако видно, что съемка велась в прохладное время года, и у многих людей на записях сложно разобрать движения из-за соответствующей одежды.

В качестве альтернативы был рассмотрен другой набор данных, состоящий из нарезки кадров из видео [4]. Он состоит из примерно 12.000 изображений, разделенных на большое количество классов по виду деятельности людей на фото, из-за чего в нем необходимо выбрать нужные для рассматриваемой задачи изображения.

После подготовки набора данных необходимо определиться с архитектурой нейронной сети. Для этого были изучены статьи со схожей тематикой [2]. В указанной работе рассмотрено определение позы человека на изображении и выделение скелета человека, что является избыточным для задачи классификации поз человека, где достаточно

только определить, бежит человек на фото или идет. В приведенной работе рекомендуется использовать сверточные нейронные сети.

Еще одна рассмотренная статья содержала обзор разных архитектур нейронных сетей, для каких классов задач они могут быть применены [1].

В настоящей работе использованы следующие подходы: язык программирования Python 3, библиотеки Keras, TensorFlow, Numpy. Для обучения были подготовлены две выборки из двух классов изображений: бегущие и идущие люди. Было написано две нейронных сети: сверточная и полносвязная. Первая нейросеть состоит из 10 слоев, она написана на Keras, а вторая нейронная сеть на TensorFlow состоит из 5 слоев. В итоге в данной работе была использована сверточная нейронная сеть, так как она показала более высокие результаты при обучении.

Материалы и методы решения

Были сформированы две обучающие выборки.

Для формирования первой обучающей выборки использованы видео, которые с помощью приложения Video to JPG Converter были преобразованы в последовательности кадров. Изначально эта выборка содержала фотографии с бегущими и идущими людьми в профиль и анфас, но из-за сложности распознавания было решено оставить только изображения, на которых люди в профиль.

Итоговая выборка состоит из 27.000 черно-белых изображений размером 160*120 пикселей. Примеры изображений приведены на рисунках 1, 2 и 3.



Рис. 1. Пример изображения с бегущим человеком.



Рис. 2. Пример изображения с бегущим человеком.



Рис. 3. Пример изображения с идущим человеком.

В этой выборке многие изображения достаточно сложно разделить по классам бегущих и идущих людей даже человеку, так как на видео были сняты не профессиональные спортсмены, что осложняет работу нейронной сети. Также изображения содержат людей с различными аксессуарами, такими как рюкзаки у мужчин (рисунок 1) и сумки у женщин.

Так как выборка получилась сложной для распознавания, изображения были обработаны, на них были выделены скелеты людей (рисунок 4).

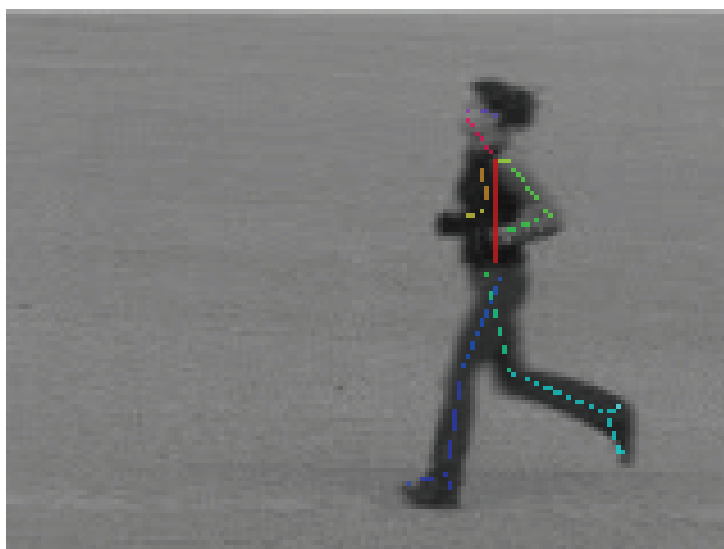


Рис. 4. Пример изображения с бегущим человеком.

Для второй выборки из 12.000 изображений были отобраны все, относящиеся к классу бегущих людей. Получилось около 400 изображений разного разрешения, пример представлен на рисунке 5.



Рис. 5. Пример изображения с бегущим человеком.

Эта выборка отличается не только разрешением изображений и их количеством, но и более разнообразными фонами на фотографиях. Также здесь съемка велась не только в профиль, и на изображениях часто больше, чем один человек, однако изображений слишком мало, чтобы можно было выбрать из них только один ракурс и фотографии с одиночными людьми.

Чтобы увеличить количество изображений в выборке, был использован обработчик для зеркального отображения фотографий. Результаты представлены на рисунке 6.

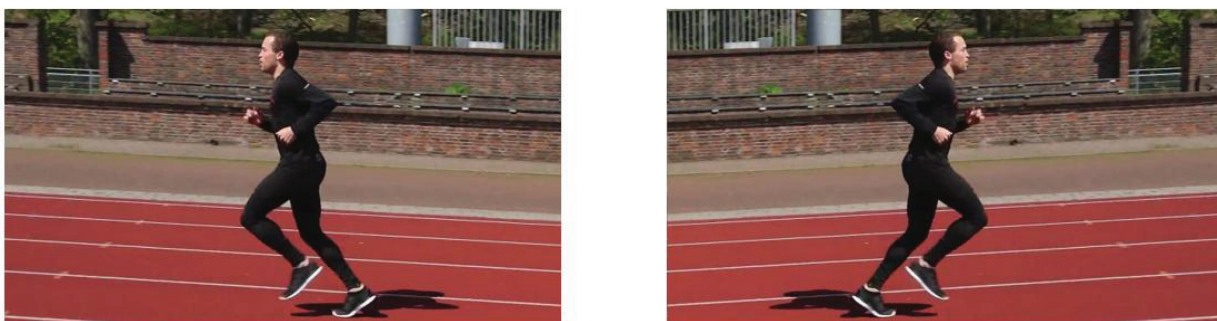


Рис. 6. Пример зеркальных изображений.

Итоговая выборка состоит примерно из 700 изображений, разделенных на бегущих и идущих людей.

Реализована сверточная нейронная сеть на языке программирования Python из 10 слоев с использованием библиотеки Keras, для оценки точности распознавания использована метрика Accuracy.

Точность классификации данной нейронной сети на первой выборке составила 50%, а на второй выборке – 75%.

Вторая полносвязная нейронная сеть реализована также на языке программирования Python, но с использованием библиотеки TensorFlow. Необходимость в ней возникла для тестирования первой обучающей выборки. Точность классификации на ней составила 50%, после чего была сформирована вторая выборка.

Результаты

Точность классификации сверточной нейронной сетью выборки из 700 изображений, разделенных на два класса, составила 75%. Нейронная сеть показала более низкие результаты на более сложной выборке, хотя в ней были подобраны фотографии, где съемка велась только в профиль и было не больше одного человека в каждом кадре.

В дальнейшем можно реализовать выделение человека на изображении и классификацию только внутри выделенной области, что поможет при классификации изображений, на которых больше, чем один человек.

Заключение

В дальнейшем можно реализовать выделение людей на изображениях, чтобы классифицировать фотографии, на которых больше, чем один человек. Также для классификации можно использовать большее число классов, реализовать классификации не по одному изображению, а по серии последовательных кадров.

Обученная модель может быть применена в различных областях, где возникает необходимость распознавать людей на изображениях и классифицировать их в зависимости от их деятельности.

Благодарности

Статья подготовлена в рамках разработки образовательного кейса для НТУ Сириус при финансовой поддержке РФФИ в рамках научного проекта № 19-37-51028.

СПИСОК ЛИТЕРАТУРЫ

1. Новые архитектуры нейронных сетей. [Электронный ресурс. <https://habr.com/ru/post/498168/>] (дата обращения: 25.05.2020).
2. Определение позы человека на изображении с помощью Forward-or-Backward Information [Электронный ресурс. <https://medium.com/@neurohiveru/определение-позы-человека-на-изображении-с-помощью-fbi-51d50c2fff88>] (дата обращения: 20.05.2020).
3. Распознавание человека по походке на основе сверточных нейронных сетей [Электронный ресурс. <https://aspirantura.hse.ru/data/2017/06/07/1170647263/2017-05-25-sokolova.pdf>] (дата обращения: 25.05.2020).
4. MPII human pose dataset [Электронный ресурс. <http://human-pose.mpi-inf.mpg.de/#download>] (дата обращения: 26.05.2020).

5. Recognition of human actions [Электронный ресурс.
<https://www.csc.kth.se/cvap/actions/>] (дата обращения: 25.05.2020).