

Нейросетевая модель трекинга маркерных объектов по видеопотоку

И.А. Килбас¹, Р.А. Парингер^{1,2}, А.В. Гайдель^{1,2}, Е.В. Гошин^{1,2}, С.В. Ровнов³

¹Самарский национальный исследовательский университет имени академика С.П. Королёва, Московское шоссе, 34а, Самара, Россия, 443086

²Институт систем обработки изображений РАН - филиал ФНИЦ «Кристаллография и фотоника» РАН, Молодогвардейская 151, Самара, Россия, 443001

³Самарский государственный медицинский университет, Арцыбушевская 171, Самара, Россия, 443001

Аннотация

Трекинг маркерных объектов по видеопотоку является важным компонентом для понимания состояния и положения объектов на видео. В настоящее время популярны модели трекинга, основанные на нейронных сетях. Несмотря на их высокую точность, данные модели нередко упускают или путают те или иные группы маркерных объектов. В данной работе предлагается нейросетевая модель, что корректирует ошибки трекинга маркерных объектов, совершаемые исходной моделью, посредством анализа их относительного положения.

Ключевые слова

Маркерные объекты, распознавание объектов, нейронные сети, свёрточные нейронные сети, распознавание позы тела человека.

1. Введение

В настоящее время крайнюю актуальность имеет задача трекинга маркерных объектов по видеопотоку. Данная проблема возникает при создании отслеживании позы тела человека [1, 2], детектировании ключевых точек лица человека [3]. Рассматриваемые в данной работе подходы осуществляют трекинг маркерных объектов посредством свёрточных нейронных сетей [4].

2. Описание нейросетевой модели трекинга

Частным случаем задачи трекинга маркерных объектов на видео потоке, представляющим особый интерес, является отслеживание позы тела человека. Маркерными объектами в данном случае являются ключевые точки на теле человека: колени, кисти и т.д. Зачастую данная задача решается с применением нейросетевых моделей [1, 2, 5].

Одна из проблем, что возникает при применении данных моделей, состоит в том, что они нередко путают стороны или не распознают отдельные группы маркерных объектов. В данной работе предлагается нейросетевая модель, что позволяет компенсировать данные ошибки, посредством регрессии более полного набора маркерных объектов с корректной ориентацией.

Для оценки качества модели в роли метрики использовалось отношение распознанных маркеров к общему числу маркеров:

$$\frac{1}{K} \sum_i^K \delta_i * 100\% \quad (1)$$

где K – общее число маркеров; δ_i равно 1, если i -ый маркер был распознан, и равно 0 в противном случае.

Данная модель увеличивает количество распознанных маркерных объектов тела человека на 10% в случаях, когда значение метрики (1) меньше среднего по тестовой выборке, а также безошибочно корректирует ориентацию (попутанные стороны) маркерных объектов. Рисунок 1 показывает сдвиг в распределении распознанных точек - распознаётся больше точек.

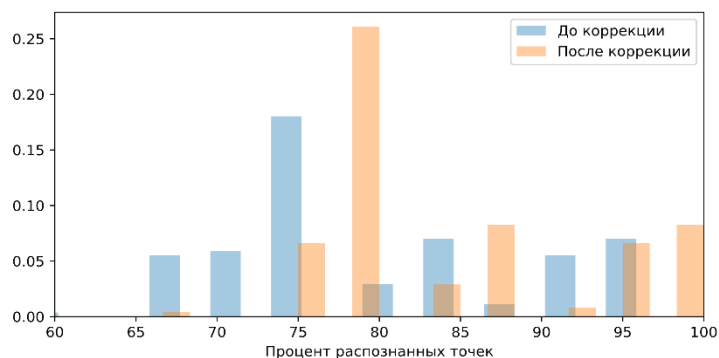


Рисунок 1: Распределение отношения количества распознанных маркеров к общему числу маркеров до и после коррекции

3. Заключение

В данной работе предложена нейросетевая модель для коррекции ошибок трекинга маркерных объектов. Данная модель способна регрессировать недостающие маркерные объекты, а также корректировать их порядок, что увеличивает итоговую точность трекинга.

4. Благодарности

Теоретические исследования выполнены при поддержке гранта РФФИ № 19-29-01235. Экспериментальные результаты получены в рамках реализации программы деятельности Лидирующего исследовательского центра, реализующего дорожную карту по "сквозной" цифровой технологии "Технологии виртуальной и дополненной реальности" при финансовой поддержке Минкомсвязи России и АО "РВК" (Договор о предоставлении гранта №003/20 от 17.03.2020 г., идентификатор соглашения о предоставлении субсидии - 0000000007119P190002).

5. Литература

- [1] Papandreou, G. Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model / G. Papandreou, T. Zhu, L.C. Chen, S. Gidaris, J. Tompson, K. Murphy // Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – P. 269-286.
- [2] Osokin, D. Real-time 2d multi-person pose estimation on CPU: Lightweight OpenPose // ArXiv preprint: 1811.12004. – 2018.
- [3] Saragih, J.M. 2011. Deformable model fitting by regularized landmark mean-shift / J.M. Saragih, S. Lucey, J.F. Cohn // International Journal of Computer Vision. – 2011. – Vol. 91(2). – P. 200-215.
- [4] LeCun, Y. Gradient-based learning applied to document recognition / Y. LeCun, L. Bottou, Y. Bengio, P. Haffner // Proceedings of the IEEE. – 1998. – Vol. 86(11). – P. 2278-2324.
- [5] Newell, A. Stacked hourglass networks for human pose estimation / A. Newell, K. Yang, J. Deng // European Conference on Computer Vision. – 2016. – P. 483-499.