**ДЕТЕКТИРОВАНИЕ И ОТСЛЕЖИВАНИЕ ИГРОКОВ НА ТРАНСЛЯЦИЯХ МАТЧЕЙ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ**

Хрящев Владимир Вячеславович

ЯрГУ им. П.Г. Демидова, г. Ярославль, Россия

1. Введение

На сегодняшний день, в области компьютерного зрения одним из ключевых направлений исследований являются алгоритмы обнаружения и отслеживания объектов на видеоданных. Под детектированием объектов понимается определение их местоположений на изображении или видеокадре. Основным отличием алгоритма отслеживания от модели обнаружения объектов является тот факт, что после обработки первого видеокадра, на котором были обнаружены целевые объекты, на последующих видеокадрах поиск прямоугольных рамок, окружающих целевые объекты, осуществляется в окрестности их местоположений [1]. И хотя человек без особого труда способен отслеживать спортсменов во время трансляций матчей, для компьютера детектирование и трекинг являются трудновыполнимыми задачами.

В данной статье представлены результаты исследования работы сверточной нейронной сети для решения задачи обнаружения и отслеживания игроков на трансляциях баскетбольных матчей. Эта статья состоит из четырех частей. Первая часть посвящена особенностям задачи детектирования и трекинга. Во втором разделе приведена архитектура сверточной нейронной сети для решения поставленной задачи. В третьей части указаны некоторые особенности обучения нейросетевой модели, а также приведены результаты ее тестирования. В заключении делаются выводы о качестве работы алгоритма глубокого обучения, а также осуществляется постановка задач для будущих исследований.

1. Алгоритм искусственного интеллекта

Задача обнаружения и отслеживания людей, в частности спортсменов, на видеоданных решалась с помощью алгоритма JDE (Joint Detection and Embedding) [2]. В качестве базовой архитектуры для него использовалась одна из разновидностей пирамидальной сети FPN (Feature Pyramid Network) – DarkNet-53 [3]. Сети FPN используют пирамидальную архитектуру для последовательного выделения признаков разного масштаба [4], что позволяет улучшить качество обнаружения объектов, размер которых варьируется на видеокадрах.

При прямом проходе входного изображения по пирамидальной сети из нее извлекаются карты признаков в трех масштабах: 1/32, 1/16, 1/8 соответственно. Далее признаки с наименьшим масштабом после применения операции повышения дискретизации объединяются с признаками большего масштаба с помощью пропускного соединения, то же самое происходит и с картами признаков большего масштаба. Далее к объединенным картам признаков всех трех масштабов добавляются небольшие сверточные сети. В конце осуществляется этап прогнозирования, при выполнении которого выделяется три набора данных:

* Результаты классификации объектов, находящихся в окантовочных прямоугольных рамках.
* Коэффициенты регрессии окантовочных прямоугольных рамок для определения на изображении их точного местоположения и формы.
* Карта встраивания объектов, для ассоциации выделенных прямоугольных рамок с траекториями движения.

На рис. 1 изображена схема архитектуры алгоритма JDE.

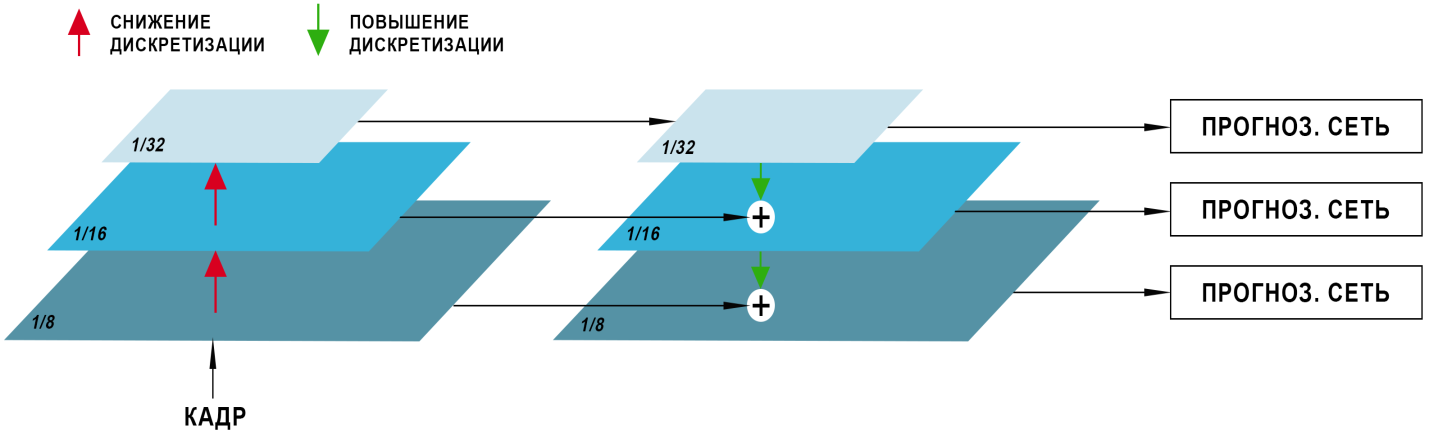


Рис. 2. – Архитектура модели JDE.

Реализация алгоритма JDE осуществлялась с помощью фреймворка PyTorch. Эта библиотека используется для решения задач компьютерного зрения и обработки естественного языка. Фреймворк PyTorch имеет большое количество встроенных модулей для предварительной обработки данных и поддержки сторонних библиотек машинного обучения [5]. В PyTorch также есть возможность выполения стандартных алгебраических и статистических операций над тензорами, единицами данных в PyTorch, как на центральном процессоре компьютера, так и на графическом ускорителе видеокарты.

1. Обучение и тестирование алгоритма

Для обучения алгоритма JDE была собрана коллекция из четырех общедоступных наборов данных по обнаружению и отслеживанию пешеходов: CalTech Pedestrian [6], CUHK-SYSU [7], PRW [8] и MOT-16 Challenge [9]. Обучающие подмножества всех этих наборов данных сформировали большую базу снимков. Во избежание переобучения алгоритма, к обучающим данным модели применялись методы аугментации: случайное вращение и масштабирование изображения, а также случайное изменение яркости, контрастности и насыщенности кадра. При необходимости, входные данные модели приводились к разрешению 1088x608 пикселей.

Тестирование алгоритма JDE осуществлялось на наборе валидационных видеофайлов с конкурса MOT-16 Challenge. Помимо этого, для проверки качества работы модели, тестирование алгоритма осуществлялось также на видеоданных со спортивных мероприятий. Для этого с помощью купольной камеры GV-EVD3100 была проведена съемка 12 отрывков с реальных игр баскетбольных команд в спортивном зале ЯрГУ им. П.Г. Демидова.

В качестве алгоритма численной оптимизации разработанной модели искусственного интеллекта использовался стохастический градиентный спуск с начальной скоростью обучения 0.01. Для более быстрой сходимости модели, на 15-ой и 23-ей эпохе обучения, скорость оптимизатора уменьшалась в 10 раз. Алгоритм JDE заканчивал свое обучение после совершения 30 эпох обучения. Критерием оценки качества алгоритма JDE служил коэффициент MOTA (Multiple Object Tracking Accuracy) [10]. Ключевой особенностью этого показателя является то, что он напрямую соотносится с тем, как человеческий глаз отслеживает целевые объекты на видеоданных. Запуск процессов обучения и тестирования алгоритмов JDE осуществлялся на графическом процессоре суперкомпьютера NVIDIA DGX-1.

В результате обучения модели и ее последующего тестирования алгоритм JDE показал результаты, сопоставимые с качеством работы современных моделей обнаружения и отслеживания объектов. Так, согласно результатам, приведенным в статьях [11, 12], алгоритм DeepSort\_2 с прогнозирующей сетью WRN (Wide Residual Connection) (DeepSort\_2 + WRN), а также модель CNNMTT с прогнозирующей пятислойной сверточной нейронной сетью (CNNMTT + 5 layer CNN), показали значения метрики MOTA, равные 61.4% и 65.2% соответственно. Для тестовой выборки базы данных MOT-16 Challenge, качество обнаружения и отслеживания объектов с помощью алгоритма JDE составило 62.6%. В то же самое время, алгоритмы DeepSort\_2 + WRN и CNNMTT + 5 layer CNN, продемонстрировали существенно более медленную скорость работы: 8 и 6 FPS соответственно против 18 FPS – скорости, полученной с помощью модели JDE. Качество работы алгоритмов детектирования и отслеживания объектов на тестовой выборке базы данных MOT-16 Challenge приведено в табл. 1.

Результаты работы алгоритма JDE для видеозаписей спортивных мероприятий, снятых с диагонального ракурса купольной камеры GV-EVD3100 показаны на рис. 2.

Рис. 2 – Результат работы алгоритма JDE для кадров видеозаписей матчей, снятых на купольную камеру.

1. Заключение

В статье были представлены результаты исследования алгоритма JDE, предназначенного для решения задачи обнаружения и отслеживания людей на видеоданных в режиме реального времени. Обучение и тестирование разработанной модели осуществлялось на графическом процессоре суперкомпьютера NVIDIA DGX-1. Численные эксперименты по оценке качества работы разработанной модели были проведены на изображениях, извлеченных из видеофайлов базы данных MOT-16 Challenge. Критерием оценки качества работы алгоритма служила метрика MOTA. Качество работы модели JDE проверялось также и на собранной базе видеозаписей баскетбольных матчей, снятых на купольную камеру GV-EVD3100, расположенную в спортивном зале ЯрГУ им. П.Г Демидова. В результате обучения и тестирования модели JDE для набора видеофайлов с конкурса MOT-16 Challenge значение метрики MOTA составило 62,6%.

В будущем планируется улучшить качество работы алгоритма JDE путем его обучения его на видеокадрах, полученных при съемке состязаний, а также с помощи более точной настройки гиперпараметров модели.

Благодарности

Исследования, приведенные в данной статье, были выполнены благодаряза поддержке федеральной программы академического лидерства «Приоритет 2030» ЯрГУ им. П.Г. Демидова, в рамках проекта «Разработка фундаментальных основ и научно-методическое обеспечение управления формированием новых компетенций для современной российской корпорации машиностроительной отрасли и формированием запросов на компетенции», финансируемого РФФИ (проект №20-010-00572А)

Список литературы

1. Минаев, Е.Ю., Кутикова, В.В., Никоноров, А.В. “ Трекинг объектов в видеопотоке на основе сверточных нейронных сетей и фрактального анализа” // Сборник трудов IV международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ-2018) - Самара: Новая техника, 2018. - С.2792-2798.
2. Wang Z., Zheng L., Liu Y., Wang S. Towards Real-Time Multi-Object Tracking. Веб: https://arxiv.org/pdf/1909.12605.pdf.
3. Redmon J., Farhadi A. YOLOv3: An Incremental Improvement. Веб: https://arxiv.org/pdf/1804.02767v1.pdf.
4. Lin T.-Y., Dollar P., Girshick R., He K., Hariharan B., Belongie S. Feature Pyramid Networks for Object Detection. Веб: https://arxiv.org/pdf/1612.03144.pdf.
5. Пойнтер Я. Программируем с PyTorch. СПб: Питер, 2020, 256 с.
6. Dollar P., Wojek C., Schiele B., Perona P. Pedestrian detection: A benchmark // 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 304-311.
7. Xiao T., Li S., Wang B., Lin L., Wang X. Joint Detection and Identiﬁcation Feature Learning for PersonSearch // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017), Honolulu, HI, USA, 2017, pp. 3376–3385.
8. Zheng L., Zhang H., Sun S., Chandraker M., Yang Y., Tian Q. Person Re-identification in the Wild. Веб:https://arxiv.org/pdf/1604.02531.pdf.
9. Milan A., Leal-Taixe L., Reid I., Roth S., Schindler K. MOT16: A Benchmark for Multi-Object Tracking. Веб: https://arxiv.org/pdf/1603.00831.pdf.
10. Щелкунов А. Е., Ковалев В. В., Морев К. И., Сидько И. В. Метрики оценки алгоритмов автоматического сопровождения // Таганрог: Известия ЮФУ. Технические науки, 2020, №1, с. 233-245.
11. Zagoruyko S., Komodakis N. Wide Residual Networks. Веб: https://arxiv.org/pdf/1605.07146.pdf.
12. Mahmoudi N., Ahadi S.M., Rahmati M. Multi-target tracking using CNN-based features: CNNMTT // Multimedia Tools and Applications, 2019, v. 78, No. 6, pp. 7077-7096.