

Использование сверточных нейронных сетей для реидентификации людей в городских условиях

Е. П. Сучков¹, Г. О. Алексеенко², К. В. Налчаджи³

В настоящее время все более широкое распространение получают различные системы видеофиксации. Одной из основных целей таких систем является контроль и слежение за человеком. Решение данной задачи позволяет в дальнейшем решать такие прикладные задачи, как контроль заполненности различных помещений (будь то торговые объекты или образовательно-культурные учреждения), построение тепловой карты перемещений человека, организация контроля доступа к тому или иному объекту. В работе предложен способ, основанный на применении комбинации различных нейронных сетей, который позволяет решать данные задачи с высокой точностью.

Ключевые слова: глубокое обучение, нейронные сети, реидентификация, Mask R-CNN, OsNet, ResNet.

Реидентификация человека является одной из основных задач в многокамерном видеонаблюдении, которое широко распространено на сегодняшний день. При решении данной проблемы подразумеваются, что изображения одного и того же человека могут быть получены с камер с неперекрывающимися ракурсами. В связи с этим, в задаче можно выделить следующие основные трудности:

- 1) Аккуратное обнаружение человека и способность отличить его от фона.
- 2) Непохожесть изображений одного и того же человека, полученных с разных ракурсов.
- 3) Сходство внешних признаков одного человека с признаками другого в местах массового скопления людей, общественных местах.

¹ *Сучков Егор Петрович* — ассистент факультета ВМК МГУ имени М.В. Ломоносова, e-mail: suchkov.egor@gmail.com.

Suchkov Egor Petrovich — assistant of CMC Faculty, Lomonosov MSU

² *Алексеенко Григорий Олегович* — главный инженер по разработке ПАО Сбербанк, e-mail: grigoriyalexeenko@gmail.com.

Alekseenko Grigoriy Olegovich — chief development engineer, PJSC Sberbank

³ *Налчаджи Карен Витальевич* — аспирант КБНЦ РАН, e-mail: nalkar07@yandex.ru.

Nalchadzhi Karen Vitalievich — graduate student, KBSC RAS



Рис. 1. Проблема 1. Выделение человека на фоне.



Рис. 2. Проблема 2. Различие изображение одного и того же человека с разных ракурсов.

В работе продолжены методы решения всех вышеописанных проблем. Для решения задачи отделения человека от фона предлагается использовать комбинацию сверточных сетей SSD-ResNet34[1] и Mask R-CNN[2], хотя допускается и применение классических подходов компьютерного зрения[3][4]. Для решения пунктов 2-3, предложено использовать сеть OsNet[5], при обучении которой применялся новый подход различных аугментаций фона. Сеть обучалась на, порядка, 30 000 уникальных персонах.



Рис. 3. Проблема 3. Сходство между разными людьми на примере 5 различных персон.

Предложенный в работе подход был протестирован различных наборах данных, включая публичные датасеты[6][7] и собственный скрытый, не участвовавший в обучении датасет.

| Название датасета \ выборки | Полученная точность | |
|--|---------------------|------------------|
| | 1 лучшая гипотеза | 5 лучших гипотез |
| WildTrack + Market-1501 (313+1501 уникальных человека), участвовала в обучении | 94% | 98% |
| Скрытый датасет (~43000 уникальных человека), не участвовала в обучении | 84% | 92% |

В сравнении с оригинальной сетью OsNet, предложенный подход позволил достичь прироста в 9% точности для сравнения по лучшей гипотезе и прироста в 5% точности для сравнения по 5 лучшим гипотезам[5], что говорит о возможности применения данного подхода и с другими реидентификационными сетями.

Список литературы

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, *Computer Vision and Pattern Recognition*, 2015.

- [2] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, “Mask R-CNN”, *Computer Vision and Pattern Recognition*, 2018.
- [3] Paul Viola, Michael Jones, “Robust Real-time Object Detection”, *International Journal of Computer Vision*, 2001.
- [4] Lucas, Bruce and Kanade, Takeo, “An Iterative Image Registration Technique with an Application to Stereo Vision (IJCAI)”, *Proceedings of Imaging Understanding Workshop*, 1981.
- [5] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, Tao Xiang, “Omni-Scale Feature Learning for Person Re-Identification”, *Computer Vision and Pattern Recognition*, 2019, <https://paperswithcode.com/paper/omni-scale-feature-learning-for-person-re/> review/?hl=5187, arXiv: <https://arxiv.org/abs/1905.00953>.
- [6] Zheng, Liang and Shen, Liyue and Tian, Lu and Wang, Shengjin and Wang, Jingdong and Tian, Qi, “Scalable Person Re-identification: A Benchmark”, *Computer Vision, IEEE International Conference on*, 2015.
- [7] T. Chavdarova; P. Baqué; A. Maksai; S. Bouquet; C. Jose et al., “WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection”, *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2018, 5030–5039.

CNN networks using for person re-identification in urban scenes Suchkov Egor, Alekseenko Grigoriy, Nalchadzhi Karen

Currently, video surveillance systems are becoming more widespread. One of the main goals of such systems is to control and track a person’s movement. The solution of this problem allows us to solve such applied problems as tracking the occupancy of various premises (whether shopping facilities or educational and cultural institutions), creating a motion heatmap or organizing control of access to a particular object. The present paper proposes a method based on a combination of various neural networks, which allows solving these problems with high accuracy.

Keywords: Deep Learning, convolutional neural network, ReID, Mask R-CNN, OsNet, ResNet.

References

- [1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”, *Computer Vision and Pattern Recognition*, 2015.
- [2] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, “Mask R-CNN”, *Computer Vision and Pattern Recognition*, 2018.
- [3] Paul Viola, Michael Jones, “Robust Real-time Object Detection”, *International Journal of Computer Vision*, 2001.
- [4] Lucas, Bruce and Kanade, Takeo, “An Iterative Image Registration Technique with an Application to Stereo Vision (IJCAI)”, *Proceedings of Imaging Understanding Workshop*, 1981.

- [5] Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro, Tao Xiang, “Omni-Scale Feature Learning for Person Re-Identification”, *Computer Vision and Pattern Recognition*, 2019, <https://paperswithcode.com/paper/omni-scale-feature-learning-for-person-re/review/?hl=5187>, arXiv: <https://arxiv.org/abs/1905.00953>.
- [6] Zheng, Liang and Shen, Liyue and Tian, Lu and Wang, Shengjin and Wang, Jingdong and Tian, Qi, “Scalable Person Re-identification: A Benchmark”, *Computer Vision, IEEE International Conference on*, 2015.
- [7] T. Chavdarova; P. Baqué; A. Maksai; S. Bouquet; C. Jose et al., “WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection”, *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2018, 5030–5039.