

ИССЛЕДОВАНИЕ ПОДХОДОВ, ОСНОВАННЫХ НА ГЛУБОКИХ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЯХ, ДЛЯ ПОВТОРНОЙ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ

Аннотация. Повторная идентификация человека — это процесс идентификации того же человека, вновь увиденного в различных полях зрения камер. Эта проблема связана с неоднозначным внешним видом человека при съемке с разных камер. Эти трудности часто усугубляются низким разрешением камер, помехами на фоне и различными условиями освещения. В последние годы был создан большой объем аннотированных наборов данных, а подходы, основанные на архитектуре глубокого обучения, за последние годы значительно улучшили точность по сравнению с ручной работой. В этой статье классифицируются подходы, основанные на глубоком обучении, на две категории: повторная идентификация человека на основе изображения и видео. Также рассмотрены работы, находящиеся в настоящее время в стадии разработки, проблемы и будущие направления для повторной идентификации личности.

Ключевые слова: повторная идентификация личности, сверточная нейронная сеть

Введение

В автоматическом многокамерном видео наблюдении под повторной идентификацией человека подразумевается, был ли один и тот же человек уже распознан в другом месте с помощью поля зрения другой камеры. Она используется для распознавания поведения, отслеживания лиц, поиска изображений и обеспечения безопасности в общественных местах. Ручной контроль системы видео не может иметь высокой точности и эффективности. Это проблема осложняется различием внешности человека на разных камерах. Таким образом, человек, наблюдаемый на изображениях со множества камер, имеет небольшие различия между классами и большие — внутри класса.

Что касается повторной идентификации личности, то уже существует несколько исследований [1–4]. В последние годы наличие большой базы аннотированных наборов данных повторной идентификации человека и успехи в глубоком обучении в области компьютерного зрения для классификации изображений и распознавания объектов также оказали большое влияние на повторную идентификацию человека. В этой статье представлены подходы, основанные на глубоком обучении, для повторной идентификации человека как в наборе изображений, так и в видеоданных.

Подходы при повторной идентификации личности, основанные на глубоком обучении в наборе изображений

В 2012 году Крижевский и соавторы представили модель глубокого обучения на основе сверточной нейронной сети [7] в конкурсе ILSVRC'12. Они выиграли это соревнование с большим отрывом в точности. С тех пор модели

глубокого обучения на основе сверточных нейронных сетей становятся все более популярными при использовании компьютерного зрения. В [5] был предложен метод глубокого обучения метрики для повторной идентификации человека с использованием двойной сверточной нейронной сети со структурой симметрии, включающей две подсети, соединенные косинусным слоем. Пара изображений используется в качестве входных данных, элементы из каждого изображения извлекаются отдельно, а затем используется их косинусное расстояние для нахождения сходства. В [6] авторами была предложена двойная архитектура, в которой используется слой patch-match, умножающий отклики сверточного элемента от двух входов на различных горизонтальных полосах и использующий произведение для вычисления сходства патчей на аналогичной широте. В [9] мягкая модель, основанная на внимании, была интегрирована с двойной нейронной сетью, чтобы адаптивно фокусироваться на важных локальных частях парных входных изображений. В [10] была представлена функция triplet loss, в которой был создан триплет из трех изображений в качестве входных данных. Каждое изображение разделено на четыре перекрывающихся части тела после первого сверточного слоя, и объединение окончательно было сделано в полностью связанном слое. В [12] авторы предложили конвейер для изучения представлений общих признаков из нескольких областей. Они объединяют все наборы данных вместе и обучают спроектированную сверточную нейронную сеть с нуля на комбинированном наборе данных, а в классификации используется функция softmax. В [13] авторы представили подход, в котором они строят один вектор Фишера [14] для каждого изображения, используя алгоритм SIFT и агрегирование цветовых гистограмм. Они использовали в качестве входных данных векторы Фишера и построили полностью подключенную сеть, а линейный дискриминационный анализ используется в качестве целевой функции.

Таблица 1. Статистика эталонных наборов изображений для повторной идентификации человека

Набор данных	Год	ID	Изображение	Камера	Метка
CUNK01 [20]	2012	971	3884	2	Ручной
CUNK02 [21]	2013	1816	7264	10	Ручной
CUNK03 [6]	2014	1360	13164	2	Ручной/DPM
PRID 450S [34]	2014	450	900	2	Ручной
Market-1501 [32]	2015	1501	32668	6	Ручной/DPM
Airport [17]	2017	9651	39902	6	ACF
DukeMTMC-reID [18]	2017	1812	36441	8	Ручной
RPIfield [19]	2018	112	601581	12	ACF

Таблица 2. Точность различных подходов глубокого обучения для повторной идентификации человека по различным наборам данных изображений (CUNK-01, CUNK-03, Market-1501, DukeMTMC-reID)

Автор/Год	CUHK-01	CUHK-03	Market-1501	DukeMTMC-reID
Li [6] (2014)	27.87%	20.65%	-	-
Wu [15] (2016)	71.14%	64.80%	37.21%	-
Xiao [12] (2016)	66.6%	75.33%	-	-
Chi-Su [11] (2016)	-	-	39.4%	-
Liu [9] (2016)	81.04%	65.65%	48.24%	-
Varior [8] (2016)	-	68.1%	65.88%	-
Wang [16] (2016)	71.80%	52.17%	-	-
Geng [22] (2016)	-	85.4%	83.7%	-
Wang [41] (2018)	-	-	98.0%	94.5%
Luo [39] (2019)	-	-	95.43%	90.2%
Quan [40] (2019)	-	-	95.4%	91.4%
Zheng [42] (2019)	-	61.1%	94.8%	90.26%

Были взяты результаты существующих подходов и можно увидеть, что преимущество в точности имеет [22] на одних из крупнейших наборах данных DukeMTMC-reID и Market-1501 (таблицы 1 и 2).

Подходы для повторной идентификации личности на основе глубокого обучения в наборе видео данных

Подходы глубокого обучения для повторной идентификации человека по наборам видеоданных представлены в [23, 25, 31], где признаки внешнего вида использовались в качестве отправной точки в РНС для получения информации о временном потоке между кадрами. В [31] была представлена структура, где сверточная нейронная сеть используется для извлечения признаков из последовательности видеок кадров и замыкается через обратную связь. В [23] авторы предложили закрытый рекуррентный блок и рекуррентную нейронную сеть на основе идентификации. В [25] и [33] предложена модель, в которой каждая входная видеопоследовательность классифицируется на соответствующий объект с использованием модели идентификации. Цветные и локальные бинарные шаблоны встроены в ячейки долгой краткосрочной памяти. В [24] предложена модель для построения гибридной сети путем объединения цветов и функций локального бинарного шаблона для извлечения пространственно-временных характеристик и внешнего вида из видеопоследовательности. В [30] авторы представили метод извлечения компактного и отличительного представления характеристик внешнего вида из выбранных кадров на основе профиля энергии потока вместо всей последовательности (таблицы 3 и 4).

Наборы данных больших размеров необходимы для обучения систем распознавания. Это большая проблема при повторной идентификации личности. Повторную идентификацию человека в открытом мире можно рассматривать как задачу проверки личности.

Таблица 3. Статистика эталонных наборов видеоданных для повторной идентификации человека

Набор данных	Год	ID	Трек	Ограничительная рамка	Камера	Метка
ETHZ [26]	2007	148	148	8580	1	Ручной
3DPES [27]	2011	200	1000	200000	8	Ручной
PRID-2011 [28]	2011	200	400	40000	2	Ручной
iLIDS-VID [29]	2014	300	600	44000	2	Ручной
MARS [33]	2016	1261	20715	1000000	6	DPM&GMMCP

Таблица 4. Точность различных подходов глубокого обучения для повторной идентификации человека по различным наборам данных (iLIDS-VID и PRIQ-2011)

Автор/Год	iLIDS-VID	PRID-2011
McLaughlin [31] (2016)	58%	70%
Zhang [30] (2017)	60.2%	83.3%
Furqan [37] (2017)	79.5%	92.5%
Pathak [38] (2019)	-	96.6%

В [35] был представлен метод для достижения низкого ложного и высокого истинного распознавания цели. В [36] предложен метод, имеющий два этапа, на первом этапе он находит, присутствует ли объект запроса в галерее или нет. На втором этапе присваивается идентификатор принятой теме запроса. Повторная идентификация людей в открытом мире все еще остается сложной задачей, о чем свидетельствует достаточно низкий уровень распознавания при низкой вероятности ложного принятия, как показано в [35,36]. Следовательно, необходимо разработать эффективные методы для повышения как точности, так и эффективности систем повторной идентификации человека.

Вывод

Повышение спроса на безопасность в общественных местах повышает интерес к повторной идентификации личности. В этой статье представлены подходы глубокого обучения как для изображений, так и для видеоданных. Решение проблемы с объемом данных, методы повторной идентификации и системы повторной идентификации в открытом мире - это важные вопросы, которые могут привлечь дальнейшее внимание сообщества.

Источники:

1. D’Orazio, T., Grazia, C.: People re-identification and tracking from multiple cameras: a review. In 19th IEEE International Conference on Image Processing (ICIP), pp. 1601–1604 (2012).

2. Bedagkar-Gala, A., Shah, S.K.: A survey of approaches and trends in person re-identification. *Image Vis. Comput.* 32(4), 270–286 (2014).
3. Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.): *Person Re-Identification. ACVPR*, vol. 1. Springer, London (2014). doi:10.1007/978-1-4471-6296-4.
4. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review. *arXiv preprint arXiv1307.5748* (2013).
5. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: *Proceedings of International Conference on Pattern Recognition*, pp. 2666–2672 (2014).
6. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid: deep filter pairing neural network for person re-identification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 152–159 (2014).
7. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: *Advances in Neural Information Processing Systems*, pp. 1097–1105 (2012).
8. Varior, R.R., Haloi, M., Wang, G.: Gated Siamese convolutional neural network architecture for human re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016. LNCS*, vol. 9912, pp. 791–808. Springer, Cham (2016). doi:10.1007/978-3-319-46484-8 48.
9. Liu, H., Feng, J., Qi, M., Jiang, J., Yan, S.: End-to-end comparative attention networks for person re-identification, *arXiv preprint arXiv:1606.04404* (2016).
10. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1335–1344 (2016).
11. Su, C., Zhang, S., Xing, J., Gao, W., Tian, Q.: Deep attributes driven multi-camera Person re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV 2016. LNCS*, vol. 9906, pp. 475–491. Springer, Cham (2016). doi:10.1007/978-3-319-46475-6 30.
12. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1249–1258 (2016).
13. Wu, L., Shen, C., van den Hengel, A.: Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification. *Pattern Recognit.* (2016).
14. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010. LNCS*, vol. 6314, pp. 143–156. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15561-1 11.
15. Wu, L., Shen, C., Hengel, A.V.D.: Personnet: person re-identification with deep convolutional neural networks. *arXiv preprint arXiv:1601.07255* (2016).
16. Wang, F., Zuo, W., Lin, L., Zhang, D., Zhang, L.: Joint learning of single-image and cross-image representations for person re-identification. In: *Proceedings of*

the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1288–1296 (2016).

17. Karanam, S., Gou, M., Wu, Z., Rates-Borras, A., Camps, O., & Radke, R. J.: IEEE Transactions on Pattern Analysis and Machine Intelligence (2018).

18. heng, Zhedong, Liang Zheng, and Yi Yang.: "Unlabeled samples generated by gan improve the person re-identification baseline in vitro." In: rXiv preprint arXiv:1701.07717 (2017).

19. Zheng, Meng and Karanam, Srikrishna and Radke, Richard J.: "RPfield: A New Dataset for Temporally Evaluating Person Re-Identification." In: CVPR Workshops (2018).

20. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012. LNCS, vol. 7724, pp. 31–44. Springer, Heidelberg (2013). doi:10.1007/978-3-642-37331-2 3.

21. Li, W., Wang, X.: Locally aligned feature transforms across views. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3594–3601 (2013).

22. Geng, M., Wang, Y., Xiang, T., Tian, Y.: Deep transfer learning for person re-identification. arXiv preprint arXiv:1611.05244 (2016).

23. Wu, L., Shen, C., Hengel, A.V.D.: Deep recurrent convolutional networks for video-based person re-identification: an end-to-end approach. arXiv preprint arXiv:1606.01609 (2016).

24. Wu, Z., Wang, X., Jiang, Y.G., Ye, H., Xue, X.: Modeling spatial-temporal clues in a hybrid deep learning framework for video classification. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 461–470 (2015).

25. Yan, Y., Ni, B., Song, Z., Ma, C., Yan, Y., Yang, X.: Person re-identification via recurrent feature aggregation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 701–716. Springer, Cham (2016). doi:10.1007/978-3-319-46466-4 42.

26. Ess, A., Leibe, B., Van Gool, L.: Depth and appearance for mobile scene analysis. In: IEEE 11th International Conference on Computer Vision, pp. 1–8 (2007)

27. Baltieri, D., Vezzani, R., Cucchiara, R.: 3DPeS: 3D people dataset for surveillance and forensics. In: Proceedings of the 2011 Joint ACM Workshop on Human Gesture and Behavior Understanding, pp. 59–64 (2011).

28. Hirzer, M., Beleznai, C., Roth, P.M., Bischof, H.: Person re-identification by descriptive and discriminative classification. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 91–102. Springer, Heidelberg (2011). doi:10.1007/978-3-642-21227-7 9.

29. Wang, T., Gong, S., Zhu, X., Wang, S.: Person re-identification by video ranking. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 688–703. Springer, Cham (2014). doi:10.1007/978-3-319-10593-2 45.

30. Zhang, W., Hu, S., Liu, K.: Learning compact appearance representation for video-based person re-identification. arXiv preprint arXiv:1702.06294 (2017).

31. McLaughlin, N., Martinez del Rincon, J., Miller, P.: Recurrent convolutional network for video-based person re-identification. In: Proceedings of

the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1325–1334 (2016).

32. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015).

33. Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., Tian, Q.: MARS: a video benchmark for large-scale person re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9910, pp. 868–884. Springer, Cham (2016). doi:10.1007/978-3-319-46466-4 52.

34. Roth, P.M., Hirzer, M., Köstinger, M., Belezni, C., Bischof, H.: Mahalanobis distance learning for person re-identification. In: Gong, S., Cristani, M., Yan, S., Loy, C.C. (eds.) Person Re-Identification. ACVPR, pp. 247–267. Springer, London (2014). doi:10.1007/978-1-4471-6296-4 12.

35. Zheng, W.S., Gong, S., Xiang, T.: Towards open-world person re-identification by one-shot group-based verification. IEEE Trans. Pattern Anal. Mach. Intell. 38(3), 591–606 (2016).

36. Liao, S., Mo, Z., Zhu, J., Hu, Y., Li, S.Z.: Open-set person re-identification. ArXiv preprint arXiv:1408.0872 (2014).

37. Furqan, M., Khan and Francois Bremond.: Multi-shot Person Re-identification using Part Appearance Mixture. (WACV, 2017).

38. Pathak, P., Erfan Eshratifar, A., Gormish, M.: Video Person Re-ID: Fantastic Techniques and Where to Find Them. arXiv preprint arXiv:1912. (2019).

39. Luo, H., Gu, Y., Liao, X., Lai, S., Jiang, W.: Bag of Tricks and A Strong Baseline for Deep Person Re-identification. arXiv:1903.07071 (2019).

40. Quan, R., Dong, X., Wu, Y., Zhu, L., Yang, Y.: Auto-ReID: Searching for a Part-Aware ConvNet for Person Re-Identification. arXiv:1903.09776 (2019).

41. Wang, G., Lai, J., Huang, P., Xie, X.: Spatial-Temporal Person Re-identification. arXiv:1812.03282 (2018).

42. Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., Kautz, J.: arXiv:1904.07223 (2019).