Person Re-identification on Mobile Devices Based on Deep Learning

Miaomiao Zhu^{a,*}, Shengrong Gong^b, Zhenjiang Qian^b, Lifeng Zhang^a

^aDepartment of Electrical and Electronic Engineering, Kyushu Institute of Technology, Kitakyushu, 8048550, Japan

^bSchool of Computer Science and Engineering, Chang Shu Institute of Technology, Changshu, 215500, China

*Corresponding Author: zhu.miaomiao234@mail.kyutech.jp

Abstract

Person re-identification, as an important supplement to face recognition, refers to a network of cameras with non-overlapping vision domains. With computer vision technology it can solve the cross-camera and cross-scene pedestrian recognition and retrieval, determining whether there is a specific pedestrian to be detected in the image or video sequence. At present, many person re-identification researches are mainly carried out through experimental verification and evaluation on large person re-identification datasets such as Market-1501, DukeMTMC-reID, MSMT17, and CUHK03. In this paper, based on the existing deep learning person re-identification research and combining with the actual application scenarios on the premise of analyzing the technical feasibility, we propose a complete process for person re-identification with mobile devices, which aims to combine pedestrian detection with person re-identification to perform real-time pedestrian detection and query. In this process, not only the features extracted by pedestrians can be reused, but also the research on person re-identification can be applied effectively, such as tracking criminals and searching for missing children.

keywords: Deep-learning, CNN, Pedestrian detection, ReID, Intelligent security, Mobile Devices.

1. Introduction

With the rapid development of computer science and technology, the research and application of artificial intelligence are gradually expanding. Every country in the world has paid much attention and support to the construction of a safe city. One of the most important parts is to install a large number of video surveillance cameras to obtain real-time information quickly and accurately in the scene. According to statistics, until September 2019, China had built the world's largest video surveillance network with more than 200 million video

lenses.⁽¹⁾

It can be seen that the installation of urban video surveillance equipment provides an effective guarantee for a safe society, especially in traffic management and criminal case investigations. However, due to the huge number of videos and images captured by cameras distributed in different regions, once a case needs to be resolved, the workload is huge if only relying on traditional ways of human to locking and analyzing the target under the multi-camera video surveillance network, to obtain information such as location, time, etc. Moreover, because of various limitations, it is easy to miss the information, which will affect the efficiency and accuracy greatly. Then, how to effectively use the face and pedestrian images captured by these cameras for real-time detection and judgment and effectively make use of it further has become an important research topic in the field of computer vision.

At present, face recognition has been effectively used in many occasions. However, in many practical application scenarios, due to the effects of the environment such as resolution, viewpoint, illumination, etc., most surveillance cameras are difficult to obtain high-quality face images, just a pedestrian image. At this time, person re-identification, as an effective supplement to face recognition, can recognize pedestrians based on information such as wear, pose, hairstyle, etc., and has become important research after face recognition technology.

At the same time, apart from the cameras in the surveillance network, mobile devices such as mobile phones are currently the main means used by people to obtain videos and images. It can be seen that face recognition based on mobile devices has been effectively applied in many occasions, such as photo album classification and real-time face authentication unlock and login in APP, and so on.

In this paper, based on the existing deep learning of person re-identification research, combining with the actual application scenarios, the paper proposes a complete process for person re-identification with mobile devices. It aims to combine

Fig. 1. Typical flow chart for person re-identification

pedestrian detection and person re-identification to perform real-time pedestrian detection and query.

2. Related work

Person re-identification(ReID), originates from multi-camera tracking and is used to determine whether the pedestrians in different images or videos captured in non-overlapping views belong to the same person. The implementation process mainly includes two parts: pedestrian detection and identification. A typical flow chart is shown in Fig. 1.

Benefiting from the development of deep learning in recent years, ReID has attracted extensive attention in many fields, such as multi-camera surveillance, multi-target tracking, public security criminal investigation, and legal evidence collection, and has good application prospects.

2.1 Deep learning

In summary, deep learning, as an implementation algorithm of machine learning, is a learning process in which deep neural networks are used for the feature representation. In other words, it can automatically generate a model by training a dataset containing a large amount of data, instead of manually extracting features by machine learning. In short, it has a strong ability to express the feature of the target and belongs to unsupervised learning. In 2006, as a representative in the field of machine learning, Geoffrey Hinton *et al.*^{(2),(3)} introduced deep learning to the field of artificial intelligence, mainly for images, speeches, natural language processing, etc., all of which are difficult to extract features.

Subsequently, Convolutional Neural Networks(CNNs), as one of the representative algorithms of deep learning, has a strong ability of representation learning (the process of convolution is also the process of feature extraction). In

1998, Yann Lecun *et al.*⁽⁴⁾ proposed LeNet-5 model and applied it to the recognition of handwritten characters. In 2012, in order to prove the potential of deep learning again, Geoffrey Hinton *et al.*, the author of^{(2),(3)} participated in *ImageNet* image recognition competition for the first time. The CNN network *AlexNet* they built had won first place twice. The accuracy rate of the network is nearly 10% higher than that of the second place. The details of *AlexNet* can refer to.⁽⁵⁾ Until then, due to its high accuracy, CNN had gradually attracted the attention of more and more researchers in computer vision and other fields. At the same time, it had gradually moved from academia to engineering, and widely used in many practical application scenarios, such as image classification, object detection, image segmentation, image annotation, image generation, etc., and had achieved great success to a great extent, including Taigman *et al.*⁽⁶⁾ landmark research achievement in the field of face recognition in 2014, Alpha Go defeating Go champion Li Shishi in 2016, and so on.

To sum up, in the field of computer vision, deep learning is the means, CNN is the method and feature learning is the purpose. In other words, a good feature representation plays a key role in the accuracy of the final algorithm.

2.2 Object detection

Object detection is to determine a specific target (object) from a static image or dynamic video sequence. While obtaining the category information, it also needs to be get its position and size, that is, Detection = Classification + Localization. Detection, classification and segmentation are the three core topics in the field of computer vision.

Before the emergence of CNN, Felzenszwalb *et al.*⁽⁷⁾ proposed DPM(Deformable parts models)detection algorithm in 2008, which was always the best algorithm in the field of object detection. DPM can be seen as an extension of HOG(Histograms of Oriented Gradients).⁽⁸⁾ Its main idea is to extract artificial features of DPM first, and then use *latent SVM (LSVM)* for classification. In practice, DPM has a good effect on general object detection, but because of its complex feature calculation and slow calculation speed, it has been limited in many application scenarios.

In recent years, with the research of deep learning, especially since the CNN achieved great success in classification tasks in 2012,⁽⁵⁾ due to its strong representation ability, deep learning has also become the best choice for image feature extraction in object detection. After that, object detection algorithms based on deep learning have been proposed successively^{(9)–(19)} and have achieved significant achievement in

Fig. 4. Number of papers on ReID that are published in top-tier computer vision conferences in recent years

Fig. 5. Sample images in the DukeMTMC-reID dataset. Note the wide range of viewpoint, obstruction, pose and illumination changes.

ber of papers on ReID that are published in top-tier computer vision conferences(CVPR, ICCV/ECCV), it can be seen that ReID has attracted more and more attention and has become a topical research area in computer vision in recent years. Especially in 2014, Wei Li *et al.*⁽³⁰⁾ introduced CNN to ReID for the first time, using the FPN network(Fiilter paring neural network) to train and getting True or False by comparing two patches. At the same time, ReID also faces several challenges. For example, compared with the sample images in the DukeMTMC-reID dataset illustrated in Fig. 5, we can see that due to the changes of camera viewing conditions, the intraclass (instance/identity) variations are typically big. For instance, some people in Fig. 5 carry a backpack, the view change across cameras(frontal to back) brings large appearance changes in the backpack area, which makes matching the same person difficult. At the same time, we must admit that the current data volume is still far from satisfactory due to the difficulties in collecting across-camera matched person images. The related researches have been done in previous papers.⁽³¹⁾

2.4 Kivy

This paper proposes research on ReID with mobile devices based on deep learning. It aims to combine the pedestrian

Fig. 6. Person re-identification based on mobile devices

detection and re-identification as described above and finally perform a real-time pedestrian detection and query through the mobile APP. Because Python has no built-in for development functions, if a complex user interface needs to be defined, one of the modules-Kivy needs to be used.

First released in 2011, Kivy is an open source Python library for rapid development of applications that makes use of innovative user interfaces, such as multi-touch APPs.⁽³²⁾ It only needs to write a set of codes and can run on major desktops and mobile platforms, including Linux, iOS, Android, Windows, and macOS. However, although Kivy is a cross-platform, if we want to run the python code on different platforms, we only need to use the packaging tool project-bulldozer for packaging python into an executable program corresponding to the platform.

3. Proposed Approach

Based on the analysis of the key technologies used above, including deep learning, object detection, and person re-identification, this section mainly introduces the overall design and partial implementation of person re-identification on mobile devices based on deep learning.

3.1 Overall design

In this paper, applying the object detection algorithm and a novel lightweight deep ReID CNN, a person-identification system based on mobile devices is built by Kivy, an open source Python library. In other words, it aims to combine pedestrian detection with person re-identification to perform real-time pedestrian detection and query. The overall implementation process of this system is shown in Fig. 6. It can be clearly seen that the system mainly includes two parts, pedestrian detection, and person re-identification.

(a) Pedestrian detection

The system receives the pedestrian images from mobile devices through real-time pedestrian detection algorithm, crop these images into a uniform size to form the candidate pedestrian library (Gallery images).

(b) Person re-identification

Use appropriate ReID models to extract the pedestrian features of the candidate image library (Gallery images) and the existing image library to be queried (Probe images), and calculate the similarity between them by distance metric. Finally, rank the results, make the candidate images with a smaller distance (or higher similarity) as query results and output them in the terminal.

Compared with the image-based ReID as shown in Fig. 3, the main difference between image-based ReID and ReID based on mobile devices is that the former is mainly carried out through experimental verification and evaluation on large existing ReID datasets, such as CUHK03, Market-1501, DukeMTMC-reID, and MSMT17. In the process of deep learning, these datasets are divided into training sets (Train) and testing sets (Query, Gallery). Pedestrian images are usually obtained by manually labeling or detection algorithm from a lot of surveillance cameras with non-overlapping fields of view. Currently, pedestrian detection and identification are independent. However, the latter's testing sets, especially the candidate pedestrian library (Gallery images), are pedestrian images captured in real time by mobile devices. detection and identification are performed simultaneously.

3.2 Techniques and methods

Based on the above analysis of system requirements and key technologies, the implementation technology proposed in this paper for person re-identification on mobile devices based on deep learning is shown in Fig. 7. The whole system is mainly composed of the following four parts.

The first part is input. In the early stage of the system design, crop the images to be retrieved and queried to a uniform size to form an image library to be queried (Probe images) as the system input.

The second part is pedestrian detection. In Section 2.2, we have described the common object detection algorithms based on deep learning in recent years. Pedestrian detection, as one of the multiple categories to be detected in the target detection task, aims to detect all pedestrians in a given image or video frame, including information such as position and size, and finally give the detection results in the form of a bounding box. Yolov3, as one of the most popular real-time

Fig. 7. Complete process for ReID on mobile devices

object detection algorithms, has high detection effects and speed. For example, it is 1000 times faster than R-CNN and 100 times faster than Fast R-CNN.⁽²²⁾ Therefore, after the comparison and analysis, we use the YOLOv3 as the initial model for pedestrian detection and pre-trained on the COCO dataset, and fine-tune the system based on the latest technology in pedestrian detection. A Keras implementation of YOLOv3⁽³⁴⁾ can detect as many as 80 categories, and we rewrite the multi-class object detection as the only category, which is "person". At the same time, in order to improve the detection speed and save the space, the images obtained by the mobile device are cropped in an interval frame manner. As shown in Fig. 8, finally, these cropped images form the candidate pedestrian library (Gallery images), the filename can contain some information such as time, location, and camera number and so on.

The third part is person re-identification. As an instance-level recognition problem, ReID relies on discriminative features, which not only capture different spatial scales but also encapsulate an arbitrary combination of multiple scales. Kaiyang Zhou *et al.*⁽³⁴⁾ proposed a novel deep ReID CNN, termed Omni-Scale Network(OSNet), that is capable of learning omni-scale feature representations. As a lightweight CNN architecture, OSNet is achieved by designing a residual block composed of multiple convolutional streams, each detecting features at a certain scale. The activation maps of OSNet and the single-scale baseline are compared in Fig.9. The visualization shows that OSNet can capture the local dis-

Fig. 8. Form the candidate pedestrian library (Gallery images) through real-time pedestrian detection

Fig. 9. Visualization of activation maps obtained by OSNet (middle one in each triplet) and the single-scale baseline (right one in each triplet)

criminative patterns of Person A (e.g., the clothing logo), which distinguishes Person A from Person B. In contrast, the single-scale model over-concentrates on the face region, which is unreliable for ReID due to the low resolution of surveillance images. Therefore, after studying and comparing multiple existing person re-identification models, we use OSNet and Torchreid,^{(35),(36)} a library for deep-learning person re-identification in PyTorch, fast end-to-end training, testing and evaluation of the deep person re-identification model.

The last part is output. This is most difficult in this system, that is, how to integrate pedestrian detection and person re-identification via Kivy. To some extent, ReID is a sorting problem. The final output results are mainly based on the common evaluation indicators: rank-k and Cumulative Match Characteristic (CMC) curve. Take the gallery images sorted in the top K as the output result on the terminal of the mobile device and set warning information based on the ratio of correct results. The early warning information is set according to the ratio of the correct results. For example, if the similarity of the detected pedestrian features after a distance measurement exceeds a certain threshold, an alarm will ring on the mobile device terminal.

3.3 Application scenarios

The purpose of this system design is to effectively combine pedestrian detection and person re-identification. In addition, make person re-identification technology more effective and practical, instead of just testing and verifying on some existing datasets, such as tracking criminals online and searching for missing children. Further designed and improved, this system can be embedded into each user's mobile phone by the public security department. Once the user shoots a specific pedestrian picture (for example, fugitives and missing children) into the lens in the usual camera process, the mobile terminal and the system background will give information prompts at the same time, including the specific location of the person involved. Because these people to be queried are iffy, and the issues of privacy and infringement are not involved, their pictures can be collected in the early stage to build a library of images to be queried (Probe images) for subsequent person re-identification. Finally, complete the online query.

4. Discussion and Future Work

For the research of this topic, the preliminary work mainly involves requirements analysis and key technology study. As mentioned above, this design mainly includes three parts: pedestrian detection, person re-identification, and Kivy. At present, pedestrian detection and APP interface design have been completed. The next part includes the following aspects: (1) Construct a person re-recognition model and extract the features of pedestrians in all images (Probe image and Gallery images); (2) Calculate the similarity between Probe image and Gallery images by distance metric; (3) Find the best match and output the real-time result based on the calculation of similarity distance; (4) Kivy, designing a person re-identification app based on mobile terminal devices for practical scenarios, such as tracking criminals and searching for missing children.

5. Conclusions

Person re-identification is an effective supplement to face recognition. Because of its important applications in security and surveillance, it has been drawing lots of attention from both academia and industry. It is not difficult to find that face recognition has played an important role in the criminal investigation. For example, in China, there are many criminals who were arrested by face recognition in 2019, such as Yuwu Xie, Zhilao Rong, Mei Li, etc. In this paper, we propose a complete process for person re-identification with mobile devices, which aims to combine pedestrian detection

with person re-identification to perform real-time pedestrian detection and query. In this process, not only the features extracted by pedestrians can be reused, but also the research on person re-identification can be applied effectively, such as tracking criminals and searching for missing children. It is worth pointing out that, benefiting from the development of deep learning in recent years, face recognition, pedestrian detection, and person re-identification will be applied to more scenarios, the effective combination of these techniques will become more powerful, and will raise the cognitive level of artificial intelligence to a new stage.

Acknowledgment

This research was supported by the Natural Science Foundation of Jiangsu Province in China under grant No. BK20191475 and Science and Technology Project of State Grid Corporation in 2019 (Research on Lightweight Active Immune Technology for Power Monitoring System).

References

- (1) https://www.sohu.com/a/342203143_100278610, access date: 2019.9.20
- (2) Geoffrey E. Hinton, Simon Osindero, and Yee-Whey Teh: "A fast learning algorithm for deep belief nets", In *Neural Computation* 18:1527-1554, Vol. 18, No. 7, pp. 1527-1554, 2006
- (3) G. E. Hinton and R. R. Salakhutdinov: "Reducing the dimensionality of data with neural networks", In *Science*, Vol. 313, No. 5786, pp. 504-507, 2006
- (4) Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner: "Gradient-based learning applied to document recognition", In *Proceedings of the IEEE* 86(11) 18:2278-2324, Vol. 86, No. 11, pp. 2278-2324, 1998
- (5) Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton: "ImageNet Classification with Deep Convolutional Neural Networks", In *Advances in Neural Information Processing Systems* 25 (*NIPS*), pp. 1097-1105, 2012
- (6) Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf: "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1701-1708, 2014
- (7) Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester, and Deva Ramanan: "Object Detection with Discriminatively Trained Part Based Models", In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 9, pp. 1627-1645, 2010
- (8) N. Dalal, and B. Triggs: "Histograms of oriented gradients for human detection", In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 886-893, 2005
- (9) Ross Girshick, J. Donahue, T. Darrell, and J. Malik: "Rich feature hierarchies for accurate object detection and semantic segmentation", In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580-587, 2014
- (10) Ross Girshick, Forrest Iandola, Trevor Darrell, and Jitendra Malik: "Deformable Part Models are Convolutional Neural Networks", In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 437-446, 2015
- (11) Ross Girshick: "Fast R-CNN", In *The IEEE International Conference on Computer Vision (ICCV)*, pp. 1440-1448, 2015
- (12) Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun: "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", In *Advances in Neural Information Processing Systems* 28 (*NIPS*), pp. 91-99, 2015
- (13) Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick: "Mask R-CNN", In *The IEEE International Conference on Computer Vision (ICCV)*, pp. 2961-2969, 2017
- (14) Tsung-Yi Lin, Piotr Dollar, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie: "Feature Pyramid Networks for Object Detection", In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2117-2125, 2017
- (15) Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi: "You Only Look Once: Unified, Real-Time Object Detection", In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779-788, 2016
- (16) Joseph Redmon, and Ali Farhadi: "YOLO9000: Better, Faster, Stronger", In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263-7271, 2017
- (17) Joseph Redmon, and Ali Farhadi: "YOLOv3: An Incremental Improvement", arXiv:1804.02767, 2018
- (18) Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg: "SSD: Single Shot Multibox Detector", In *European Conference on Computer Vision (ECCV)*, Vol. 9905, pp. 21-37, 2016
- (19) Golnaz Ghiasi, Tsung-Yi Lin, Ruoming Pang, and Quoc V. Le: "NAS-FPN: Learning Scalable Feature Pyramid Architecture for Object Detection", In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7036-7045, 2019

- (20) Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato and Lior Wolf: "DeepFace: Closing the Gap to Human-Level Performance in Face Verification", In The IEEE Conference on Computer Vision and Pattern Recognition (*CVPR*) , pp. 1701-1708 , 2014
- (21) J. R. R. Uijlings, K.E.A. van de Sande, T. Gevers, and A. W. M. Smeulders: "Selective Search for Object Recognition", In International Journal of Computer Vision, Vol. 104, pp. 154-171, 2013
- (22) Xiongwei Wua, Doyen Sahoo and Steven C. H. Hoi: "Recent Advances in Deep Learning for Object Detection", In Neurocomputing, arXiv:1908.03673, 2019
- (23) Wojciech Zajdel, Zoran Zivkovic, and Ben J. A. Krose: "Keeping Track of Humans: Have I Seen This Person Before?", In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, DOI: 10.1109/ROBOT.2005.1570420, 2005
- (24) Douglas Gray, Shane Brennan, and Hai Tao: "Evaluating appearance models for recognition, reacquisition, and tracking", In Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (*PETS*) , vol. 3, No. 5, 2007
- (25) Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang: "DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification", In The IEEE Conference on Computer Vision and Pattern Recognition (*CVPR*) , pp. 152-159, 2014
- (26) Chen Change Loy, Tao Xiang, and Shaogang Gong: "Multicamera activity correlation analysis", In 2009 IEEE Conference on Computer Vision and Pattern Recognition (*CVPR*) , 2009
- (27) Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian: "Scalable person re-identification: A benchmark", In The IEEE International Conference on Computer Vision (*ICCV*) , pp. 1116-1124, 2015
- (28) Zhedong Zheng, Liang Zheng, and Yi Yang: "Unlabeled samples generated by gan improve the person re-identification baseline in vitro", In The IEEE International Conference on Computer Vision (*ICCV*) , pp. 3754-3762, 2017
- (29) Longhui Wei, Shiliang Zhang, Wen Gao, and Qi Tian: "Person Transfer GAN to Bridge Domain Gap for Person Re-Identification", In The IEEE Conference on Computer Vision and Pattern Recognition (*CVPR*) , pp. 79-88, 2018
- (30) Wei Li, Rui Zhao, Tong Xiao, and Xiaogang Wang: "DeepReID: Deep Filter Pairing Neural Network for Person Re-Identification", In The IEEE Conference on Computer Vision and Pattern Recognition (*CVPR*) , pp. 152-159, 2014
- (31) Miaomiao Zhu, Shengrong Gong, Zhenjiang Qian, and Lifeng Zhang: "A Brief Review on Cycle Generative Adversarial Networks", In The 7th IIAE International Conference on Intelligent Systems and Image Processings (*ICISIP*) , pp. 235-242, 2019
- (32) <https://kivy.org/doc/stable/>, access date: 2019.7.20
- (33) <https://github.com/qqwweee/keras-yolo3/>, access date: 2019.9.21
- (34) Kaiyang Zhou, Yongxin Yang, Andrea Cavallaro and Tao Xiang: "Omni-Scale Feature Learning for Person Re-Identification", In The IEEE International Conference on Computer Vision (*ICCV*) , pp. 3702-3712, 2019
- (35) Kaiyang Zhou, and Tao Xiang: "Torchreid: A library for Deep Learning Person Re-Identification in Pytorch", arXiv:1910.10093, 2019
- (36) <https://zhuanlan.zhihu.com/p/28342644>, access date: 2019.5.10