

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Person Re-Identification Using Deep Learning Approach

Mr. Saravanan K¹, Jeevitha R², Keerthana R³, Yogapriya K⁴

Assistant Professor, Department of Information Technology, R.M.D Engineering College, Kavaraipettai,
Chennai, India¹

Student, Department of Computer Engineering, R.M.D Engineering College, Kavaraipettai, Chennai, India²

ABSTRACT: Re-identification of a person (Re-ID) has gained significant attention because of its ability to recognize a person of interest from various surveillance images. High computing and storage costs have presented a major challenge for resource-constrained users with a growing number of surveillance videos. However, individual re-ID may result in a security threat over outsourced surveillance videos, i.e., the innocent person's privacy leakage in those videos. The proposed solution is, therefore, effective privacy- individual re- over outsourced surveillance videos using the Haar cascade algorithm, which can guarantee the identified person's privacy while providing re- of the individual. In particular, deep learning and then extracting successful individual re-identification features would determine whether or not the individual faces re-identified.

KEYWORDS: Deep learning, Face detection

I. INTRODUCTION

Individual re-identification is aimed at recognizing or re-identifying an individual previously seen in a video recorded with the overlapping or disjoining field of view. It has many applications in video surveillance, such as monitoring people in a crowd as they pass through the field of view of different cameras or in and out of a field of view of the same camera. Re-identification helps us to reduce the likelihood of identity switches by manipulating a person's appearance to re-identify him/her again and it has been expanded like an individual re-identification to the question of tracking multiple players in sports videos whose paths can overlap and trigger identity switches. This is a challenging problem due to occlusions and changes in illumination and poses caused by a change in viewing angle and lighting conditions when the individual moves in a scene or from one field of view to another, and uses appearance symbols such as color histograms, jersey numbers and facial recognition to minimize identity switches.

Deep learning is the secret to voice control in electronic devices such as phones, laptops, televisions, and hands-free orators. Lately and with good cause, deep learning is getting lots of attention. It's producing results which were previously not possible. In deep learning a computer, the model learns directly from pictures, text, or sound to perform classification tasks. Deep learning models can achieve cutting-edge precision, often exceeding output at the human level. To get the right answer, it needs to define the actual problem and it should be recognized that the Deep Learning feasibility should also be tested (whether it would suit Deep Learning or not). This needs to define the appropriate data that will lead to the real issue and plan accordingly. Use the right Algorithm for Deep Learning.

The problem of matching people through disjoint camera views in a multi-camera system is known as the person re-identification face (or person re-id for short). This is useful for a range of public safety applications, such as smart camera surveillance. In a standard real-world application, a single person faces, or a watch list of several known individuals is given as the target set for searching through a wide amount of video surveillance footage where the

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

persons on the watch list are likely to reappear. This project aims to apply a person re-identification algorithm, such as Haar Cascade, which can tell which person is identical to the face of an individual already observed and which face is new or dissimilar to all faces observed to obtain satisfactory results on different datasets without any training.

II. RELATED WORK

To compile a VAC21 vehicle dataset and carefully mark 21 groups of attributes to the structure. We apply the state-of-the-art one-stage detection method SSD as a reference model for detection of attributes with the built dataset. To make improvements by introducing low-level layer proposals to increase the precision of detection of small artifacts. In addition, we use the focal loss to boost detection accuracy, then we suggest a new vehicle re-identification and retrieval system based on ROIs, and perform a series of experiments. The experimental findings indicate that in literature, our approach outperforms state-of-the-art methods. Re-identification of the vehicles plays a significant role in applications for video surveillance. Due to numerous factors such as pose variability, shifts in lighting, and subtle inter-disparities, it remains a daunting job given the efforts made on this issue in the last few years. It believes the primary identifying knowledge in the literature has not been well explored. To collect a 'VAC21' vehicle dataset which contains 7,129 images of five vehicle types? Then, the 21 groups of structural attributes are carefully numbered hierarchically with bounding boxes. It knows the dataset which is labelled with many comprehensive attributes. Subsequently, to make a few significant improvements specific to this application to increase accuracy by introducing more low-level layer proposals to improve the accuracy of detecting small objects and using focal loss to improve mean average accuracy. In addition, the attribute detection results can be applied to a series of vision tasks that concentrate on the interpretation of the vehicle images. Such profound features are input into a boosting model to increase accuracy. A series of tests is performed on the Vehicle dataset and the state-of-the-art techniques are the experimental outcomes. Discussing the drawbacks of this method is that it has not done growth patterns. It has not carried out a systematic analysis of smart contracts and recognizes certain weaknesses in the study. Some other important works are

- An Optimization Process for Enhancing the Cascade of People Detector Classifiers.
- A New SIFT Feature Detection and Object Tracking Selection System.
- A hybrid particulate weighting method for SIFT particulate filtering functionality.
- A System for Object Tracking on Software Association.
- Easy System for Managing Occlusion.

III. LITERATURE SURVEY

Title: 3-D Person VLAD: Learning Deep Global Representations for Video-Based Person Re-identification

Author: Lin Wu, Yang Wang, Ling Shao, and Meng Wang

Year: 2019

It presents end-to-end trainable deep neural networks to produce global video-level features over the entire video span for video-based person re-ID. The major contribution is to build up a deep 3-D convolution architecture with an amenable layer, namely, a 3-D PersonVLAD aggregation layer, to encode spatiotemporal signals into a global compact video descriptor. To make the learned video features discriminative against various misalignments, a 3-D part alignment module based on the attention mechanism is introduced into the feature learning stage to localize distinct regions from which local features are aggregated to yield the global robust representations. With the global space-time convolutions over full-length video extent, it achieves state-of-the-art performance on three benchmark data sets. It also demonstrates the generalization and scalability of our model. It presents the global deep video representation learning to video-based person re-identification (re-ID) that aggregates local 3-D features across the entire video extent. Existing methods typically extract frame-wise deep features from 2-D convolutional networks (Convents) which are pooled

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

temporally to produce the video-level representations. However, 2-D ConvNets lose temporal priors immediately after the convolutions and a separate temporal pooling is limited in capturing human motion in short sequences. It presents global video representation learning, to be complementary to 3-D ConvNets as a novel layer to capture the appearance and motion dynamics in full-length videos. Nevertheless, encoding each video frame in its entirety and computing aggregate global representations across all frames is tremendously challenging due to the occlusions and misalignments. To resolve this, our proposed network is further augmented with the 3-D part alignment to learn local features through the soft-attention module. These attended features are statistically aggregated to yield identity-discriminative representations. Our global 3-D features are demonstrated to achieve state-of-the-art results on three benchmark data sets: MARS, Imagery Library for Intelligent Detection Systems-Video Re-identification, and PRID2011.

Title: Discriminative Feature Learning with Foreground Attention for Person Re-identification

Author: Sanping Zhou, Jinjun Wang, DeyuMeng, Yudong Liang, Yihong Gong, Nanning Zheng

Year: 2019

It proposes a simple yet effective deep neural network to learn a discriminative feature representation from the foreground of each input image for person Re-ID. Firstly, a FANN is constructed to jointly enhance the positive influences of foregrounds and weaken the side effects of backgrounds, in which an encoder and decoder network is built to guide the whole network to directly learn a discriminative feature representation from the foreground persons. Secondly, a novel local regression loss function is designed to deal with the isolated regions in the ground truth masks by considering the local information in a neighborhood. Thirdly, the asymmetric triplet loss function is introduced to supervise the feature learning process, which can jointly minimize the intra-class distance and maximize the inter-class distance in each triplet unit. Extensive experiments on the 3DPeS, VIPeR, CUHK01, CUHK03, Market1501 and DukeMTMC-Reid datasets are conducted, and the results have shown that our method can significantly outperform the state-of-the-art approaches. The performance of person re-identification (Re-ID) has been seriously affected by the large cross-view appearance variations caused by mutual occlusions and background clutters. Hence learning a feature representation that can adaptively emphasize the foreground persons becomes very critical to solving the person Re-ID problem. In this paper, we propose a simple yet effective foreground attentive neural network (FANN) to learn a discriminative feature representation for person Re-ID, which can adaptively enhance the positive side of the foreground and weaken the negative side of the background. Specifically, a novel foreground attentive subnetwork is designed to drive the network's attention, in which a decoder network is used to reconstruct the binary mask by using a novel local regression loss function, and an encoder network is regularized by the decoder network to focus its attention on the foreground persons. The resulting feature maps of the encoder network are further fed into the body part subnetwork and feature fusion subnetwork to learn discriminative features. Besides, a novel symmetric triplet loss function is introduced to supervise feature learning, in which the intra-class distance is minimized and the inter-class distance is maximized in each triplet unit, simultaneously. Training our FANN in a multi-task learning framework, a discriminative feature representation can be learned to find out the matched reference to each probe among various candidates in the gallery. Extensive experimental results on several public benchmark datasets are evaluated, which have shown clear improvements of our method over the state-of-the-art approaches.

Title: Fuzzy Multilayer Clustering and Fuzzy Label Regularization for Unsupervised Person Re-identification

Author: Zhong Zhang, Meiyan Huang, Shuang Liu, Baihua Xiao, and Tariq S. Durrani

Year: 2019

It has proposed a novel method termed FMC for unsupervised person re-identification. The proposed FMC possesses two advantages. Firstly, to alleviate the influence of complex pedestrian images, FMC utilizes a multilayer perceptron to map features into a new feature space which is beneficial for clustering. Secondly, FMC assigns fuzzy labels for unlabeled pedestrian images instead of onehot label, which could simultaneously consider the membership degree and the similarity between the sample and each cluster. Afterward, the proposed FLR utilizes pedestrian images with fuzzy labels to train ResNet-50, which regularizes ResNet-50 learning process and meanwhile reduces the risk of over-fitting.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Experiments on three benchmark person re-identification databases demonstrate that our method achieves a new state of the art. Unsupervised Person re-identification has received more attention due to its wide real-world applications. In this paper, we propose a novel method named fuzzy multilayer clustering (FMC) for unsupervised person re-identification. The proposed FMC learns a new feature space using a multilayer perceptron for clustering to overcome the influence of complex pedestrian images. Meanwhile, the proposed FMC generates fuzzy labels for unlabeled pedestrian images, which simultaneously considers the membership degree and the similarity between the sample and each cluster. We further propose the fuzzy label regularization (FLR) to train the convolutional neural network (CNN) using pedestrian images with fuzzy labels in a supervised manner. The proposed FLR could regularize the CNN training process and reduce the risk of over-fitting. The effectiveness of our method is validated on three large-scale person re-identification databases, i.e., Market-1501, DukeMTMC-re-ID and CUHK03.

IV. PROPOSED SYSTEM

Due to its application and research importance, the re-identification of persons (re-ID) has become increasingly common within the community. It spots a person of interest in other cameras. This has seen the advent of large-scale data sets and deep learning systems that use large volumes of data. Considering we have to find motion detection because if anyone crosses the camera it first recognized motion then face after that face identification. We have already store image dataset including projects. The camera identifies whether the person has already present or not by using that dataset. The image data use is usually split into training images from a given input video file. The training set contains a known output and the model learns on these images to be generalized to another image later on. It has the test video to test our models and it will do this using the Tensor flow library in Python using the Keras method by haar cascade algorithm.

There are 5 modules in person re-identification application

- Person face recognition from a given input image
- To capture the faces from the given video for training purpose
- To build a model for the training process
- Implementation of person faces re-identification by Haar cascade classifier
- GUI application

PERSON FACE RECOGNITION FROM INPUT IMAGE:

The face can be assumed to be an individual's unique identity. Individuals worldwide have distinctive features and facial characteristics. This plays a significant role in society's relationship with other men. In the real world, facial recognition is applied in light of this evidence. A Haar Wavelet uses box-shaped patterns to detect signals with abrupt transformations as it produces square-shaped waves. A scale is chosen smaller than the target image for the study of an image using Haar cascades. This is then placed on the image, and the pixel values in each segment are taken on average. If a given threshold passes the difference between two values it is called a match. Face detection on a human face is achieved by matching various haar-like characteristics.

TO CAPTURE THE FACES FROM GIVEN VIDEO FOR TRAINING PURPOSE:

OpenCV allows XML files to be generated to store functions derived from datasets using the Face Recognizer class. The stored images are imported, converted to grayscale and saved with IDs in two lists with the same indexes. Face Recognizer objects are created using face recognizer class. A face recognizer object is created using the desired parameters. The facial detector is used for the identification, cropping, and transition of faces to be identified in the image. This is done using the same technique used for the image capture application. This is achieved using the same

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

technique as for collecting pictures. A prediction is made for each face detected using FaceRecognizer.predict() which returns the class ID and confidence.

TO BUILD A MODEL FOR TRAINING PROCESS:

Cascade training is not a simple work. The first important bit was to collect the images, then create samples based on them, and finally start the workout. The OpenCV train cascade utility is an improvement in many respects over its predecessor, one of which is that train cascade allows multithreading of the training process which reduces the time it takes to complete the classifier training. This multithreaded approach is only applied during the pre-calculation step however, so the overall time to train is still quite significant, resulting in hours, days and weeks of training time. Because the training process involves a lot of positive and negative feedback images, which might not always be available, so using a method for generating these positive images is a way to overcome this. OpenCV built-in mode allows more positive images to be generated with distortion of the original positive image and application of a background image. However, it does not allow us to do this for multiple images. By using the Perl script create samples to apply distortions in batch and the merge vector tool, it is possible to create such necessary files for each positive input file and then merging the outputted files into one input file that OpenCV can understand. The number of positives and negatives is another significant thing to remember. To begin training, the number of positive and negative images must be entered while executing the instruction.

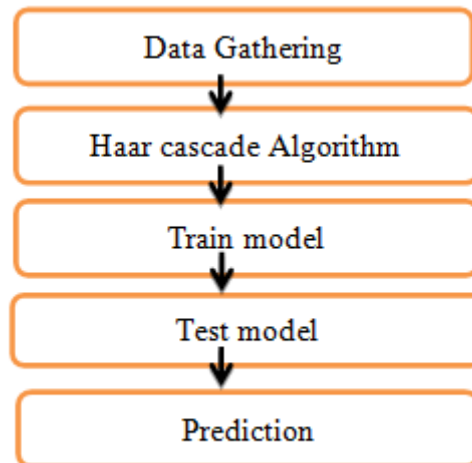


Fig.1: Steps of process by dataflow diagram

IMPLEMENTATION OF PERSON FACES RE-IDENTIFICATION BY HAAR CASCADE CLASSIFIER:

Haar feature-based cascade classifiers are efficient object detection, and it is a machine-based learning technique that trains a cascade function from a lot of positive and negative images. It will work with face detection, the algorithm needs a lot of positive images (images of faces) and negative images (images without faces) to train the classifier. We then have to draw characteristics from it. That feature is a single value obtained by subtracting the number of pixels from the number of pixels under a black rectangle under a white rectangle. The cascade classifier is a sequence of stages in which each stage is a sequence of poor learners. The system detects a face in question by moving a window over the image. Each stage of the classifier labels the specific region defined by the current location of the window as either positive or negative – positive meaning that a face was found or negative means that the specified object was not found in the image. If the labeling yields a negative result, then the classification of this specific region is hereby complete and the location of the window is moved to the next location. If the labeling gives a positive result, then the region moves off to the next stage of classification.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

GUI APPLICATION:

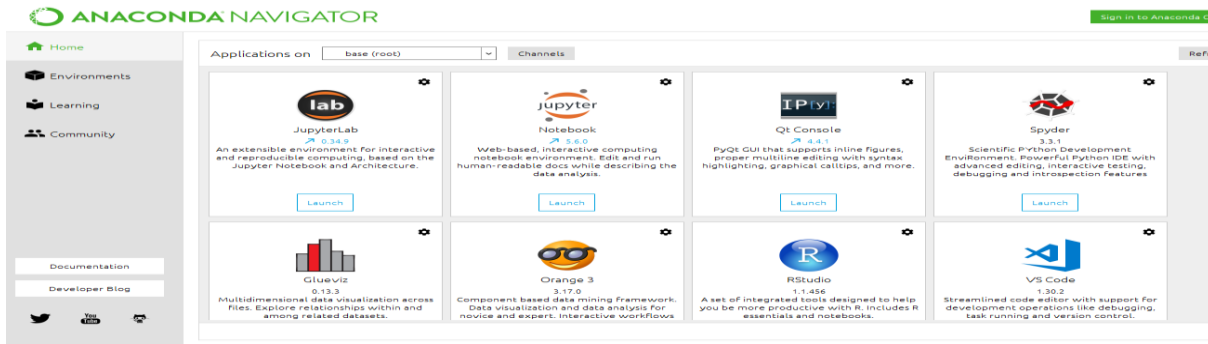


Fig.2: Open Anaconda navigator

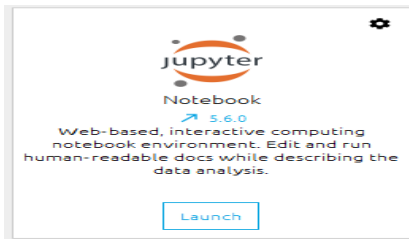


Fig.3: Launch the Jupyter notebook platform

V. EXPERIMENTAL RESULTS

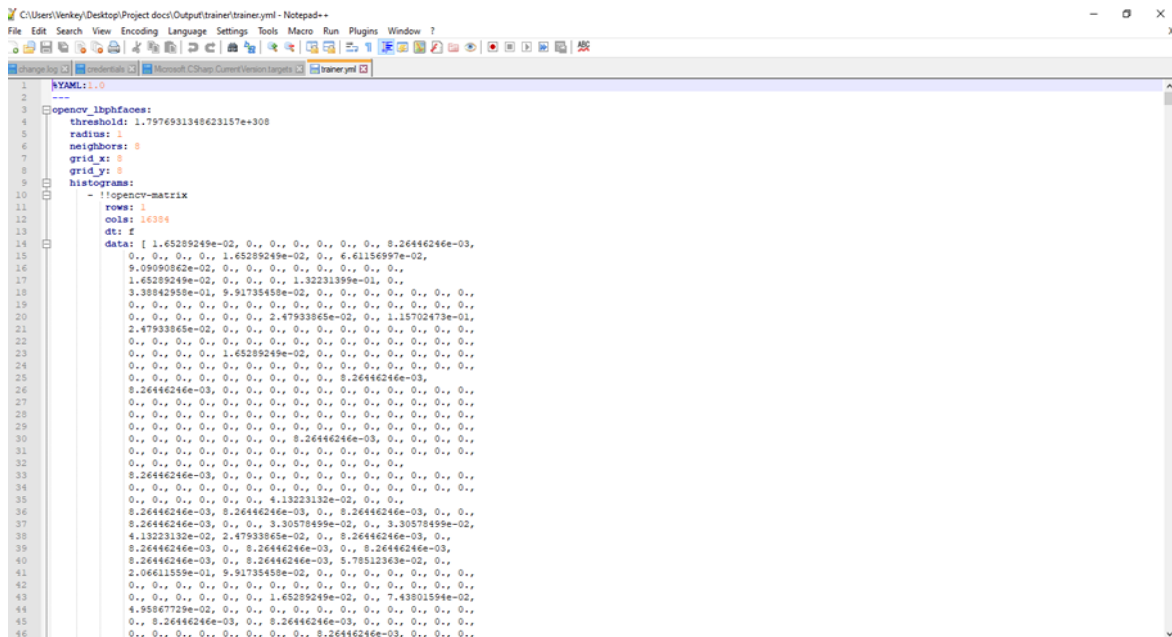


Fig.4: Datasets of the image in YML format

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

Test No	Module name	Actual input	Expected output	Actual output	Pass/Fail
T01	Detection of the human face	Giving a video source camera 1 as input	The face is expected to be detected with a blue box shape	The face is detected with a blue box shape	Pass
T02	Detection of eyes	To detect the eyes from the given video source	The eyes are expected to be detected with a blue box shape	The eyes are detected with a blue box shape	Pass
T03	Obtaining datasets	To obtain datasets as images from given video source cam1.mp4	The images are stored with ID under datasets folder	Images are extracted and stored in datasets folder	Pass
T04	Training datasets	The faces are trained by the trainer and stored as YML file	The datasets are expected to be trained using a trainer	Datasets are trained and stored as YML file	Pass
T05	Person re-identification	The person trained is detected from another video source cam2.mp4	The person is expected to be detected from a given video	The person is detected with name.	Pass

The below table shows the input and corresponding output.

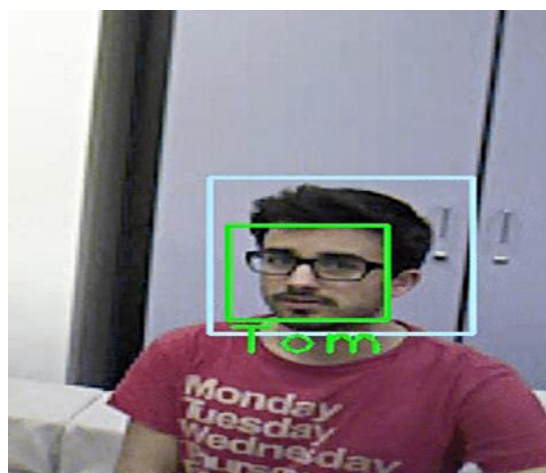


Fig.5: Identified person indicated with name and green box

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 9, Issue 3, March 2020

VI. CONCLUSION

The Haar cascade algorithm is implemented to classify human and we addressed the task of facial expression recognition by identified face image through the input video file. It classified the image of faces into any of two discrete face features categories that represent end-to-end trainable deep learning algorithms to produce global video-level features over the entire video for video-based person re-ID.

REFERENCES

- [1] Y.-G. Jiang, Z. Wu, J. Tang, Z. Li, X. Xue, and S.-F. Chang, "Modelling multimodal clues in a hybrid deep learning framework for video classification," *IEEE Trans. Multimedia*, 2018.
- [2] D. Zapletal and A. Herout, "Vehicle re-identification for automatic video traffic surveillance," in *Proc. IEEE Conf. Computing. Vis. Pattern Recognition. Workshops (CVPRW)*, Jun. 2016, pp. 1568–1574.
- [3] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Computing. Vis. Pattern Recognition. (CVPR)*, Jun. 2015, pp. 815–823.
- [4] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognition.*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015.
- [5] W. Lin *et al.*, "Learning correspondence structures for person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2438–2453, May 2017.
- [6] Mang Ye, Jiawei Li, Andy J Ma, Liang Zheng, and Pong C.Yuen, "Dynamic graph co-matching for unsupervised video-based person re-identification," *IEEE Transactions on Image Processing (TIP)*, 2019.
- [7] L. Wu, Y. Wang, and L. Shao, "Cycle-consistent deep generative hashing for cross-modal retrieval," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1602–1612, Apr. 2019.
- [8] D. Zapletal and A. Herout, "Vehicle re-identification for automatic video traffic surveillance," in *Proc. IEEE Conf. Computing. Vis. Pattern Recognition. Workshops (CVPRW)*, Jun. 2016, pp. 1568–1574.
- [9] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet- A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Computing. Vis. Pattern Recognition. (CVPR)*, Jun. 2015, pp. 815–823.
- [10] R. Quanet *et al.*, "Unsupervised salient object detection via inferring from imperfect saliency models," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1101–1112, May 2018.
- [11] S. Ding, L. Lin, G. Wang, and H. Chao, "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognition*, vol. 48, no. 10, pp. 2993–3003, Oct. 2015.
- [12] W. Lin *et al.*, "Learning correspondence structures for person re-identification," *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2438–2453, May 2017.
- [13] H. Liu, Y. Tian, Y. Yang, L. Pang, and T. Huang, "Deep relative distance Learning, Tell the difference between similar vehicles," in *Proc. IEEE Conf. Computer. Vis. Pattern Recognition. (CVPR)*, Jun. 2016, pp. 2167–2175.
- [14] C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Multi-type attributes driven multi-camera person re-identification," *Pattern Recognition.*, vol. 75, pp. 77–89, 2018.