

REVIEW ON DEEP LEARNING BASED TECHNIQUES FOR PERSON RE-IDENTIFICATION

Abhinav Parkhi

Research Scholar, Department of Electronics & Telecommunication Engineering, YCCE, Nagpur, (India).

E-mail: abhinav.parkhi@gmail.com

Atish Khobragade

Professor, Department of Electronics Engineering, YCCE, Nagpur, (India).

E-mail: atish_khobragade@rediffmail.com

Reception: 24/11/2022 **Acceptance:** 09/12/2022 **Publication:** 29/12/2022

Suggested citation:

Parkhi, A., y Khobragade, A. (2022). Review on deep learning based techniques for person re-identification. *3C TIC. Cuadernos de desarrollo aplicados a las TIC*, 11(2), 208-223. <https://doi.org/10.17993/3ctic.2022.112.208-223>



<https://doi.org/10.17993/3ctic.2022.112.208-223>

ABSTRACT

In-depth study has recently been concentrated on human re-identification, which is a crucial component of automated video surveillance. Re-identification is the act of identifying someone in photos or videos acquired from other cameras after they have already been recognized in an image or video from one camera. Re-identification, which involves generating consistent labelling between several cameras, or even just one camera, is required to reconnect missing or interrupted tracks. In addition to surveillance, it may be used in forensics, multimedia, and robotics. Re-identification of the person is a difficult problem since their look fluctuates across many cameras with visual ambiguity and spatiotemporal uncertainty. These issues can be largely caused by inadequate video feeds or low-resolution photos that are full of unnecessary facts and prevent re-identification. The geographical or temporal restrictions of the challenge are difficult to capture. The computer vision research community has given the problem a lot of attention because of how widely used and valuable it is. In this article, we look at the issue of human re-identification and discuss some viable approaches.

KEYWORDS

Person re-identification, Supervised Learning, Unsupervised Learning.

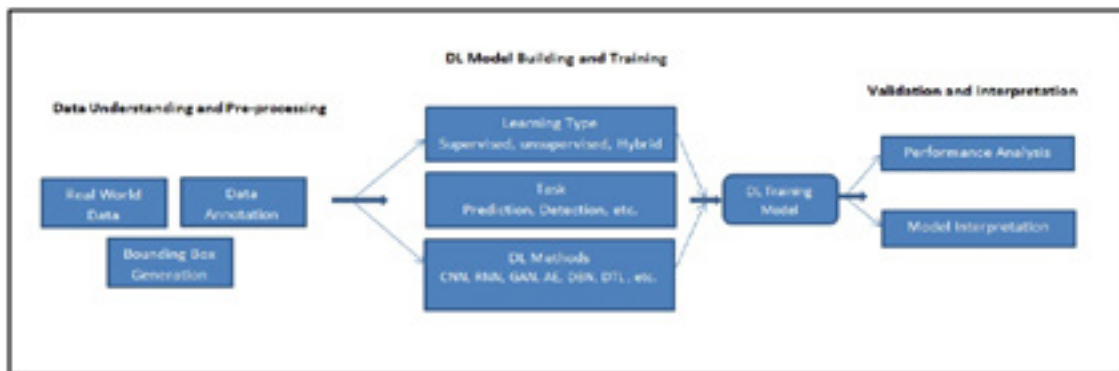
1. INTRODUCTION

The process of Person Re-identification(Re-ID) has been thoroughly studied as a distinct person retrieval problem among non-overlapping cameras [1]. Re-ID's goal is to determine whether a person of interest has ever been at a location at a different time that was captured by the same camera at a different time instant or even the same camera placed somewhere else [2]. A photograph [3], a video clip [4], or even a written explanation [5] might be used to illustrate the subject. Person Re-ID is essential in smart surveillance technology with substantial academic effect and practical advantage due to the pressing need for community security and the growing number of security cameras.

The procedure of re-ID is difficult because of a variety of camera motions [6], poor picture resolutions [7, 8, 9, 10, 11], heterogeneous modalities [11, 12], complicated camera surroundings, background clutter [12], inaccurate bounding box creation, etc. These provide a lot of variations and uncertainty. Other elements that significantly increase the difficulties for realistic model deployment include the dynamically network of upgraded cameras [13], a massive gallery that offers effective restoration [14], group ambiguity [15], important domain change [16], unknown examining situations [17], and updating a model progressively [18], and changing clothes [19]. Re-ID still presents a problem as a result of these issues. This encourages us to carry out an extensive survey, establish a solid baseline for various Re-ID efforts,

and discuss a wide range of potential future paths. Although Re-ID is a difficult process, enhancing the semantic integrity of the analysis depends on it. Re-ID is crucial for programs that make use of single-camera surveillance systems. For instance, to find out if a person regularly visits the same place or if a different person or the same one picks up an abandoned box or bag. In addition to tracking, it has uses in robotics, multimedia, and more well-known technologies like automatic photo labeling and photo surfing [20]. It is not difficult to comprehend the Person's Re-ID process. Being human, we always do it with ease. Our sights and minds have been conditioned to locate, identify, and then re-identify things and people in the actual world. Re-ID, which can be shown in Fig. 1, is the idea that a person who has been earlier seen would be identified as soon as they make an appearance using a specific description of the individual.

Even if hand-crafted features had some early success [21] and metric learning [22], the most advanced Re-ID algorithms currently available are constructed using convolutional neural networks (CNNs), which, when trained under supervision, need a significant amount of annotated (labelled) data to learn a stable embedding subspace. Recent deep learning approaches and detailed investigations on person Re-ID utilizing custom systems provided in [23], respectively. Large-scale dataset annotation for Re-ID is exceedingly labor intensive, time-consuming, and expensive, especially for techniques needing numerous bounding boxes for each individual to increase accuracy by making generalisations between two separate activities. One-shot learning and unsupervised learning are combined, for instance, in [24] and, which employ the Resnet50 [25] architecture with pre-trained networks on ImageNet [26]. Although it has been empirically demonstrated that pre-training and transfer learning significantly boost neural network performance, they are not appropriate for adjusting parameters across a wide range of domains or topologies. This article highlights the obstacles and unresolved problems in human re-identification and datasets, deep learning algorithms, and current research in these areas.



Figure

1. Shows an example of a common DL workflow that involves following five stages (i) Data Collection (ii) Bounding Box Generation (iii) Data Annotation (iv) Model Training (v) Validation [67].

2. DEEP LEARNING MODERN RESEARCH

In today's Era, intelligent systems and tech sophisticated automation are the main focuses across a diverse range of domains, including smart cities, e-Health, enterprise intelligence, innovative treatment, cyber security intellectual ability, and many more [27]. Particularly when it comes to security technologies as a wonderful approach to disclose complicated data structures in high dimensions, deep learning techniques have substantially improved in terms of effectiveness across a wide range of applications. In order to create intelligent data-driven systems that satisfy current expectations, DL techniques might be extremely important because to their exceptional learning capabilities from past data. DL has the ability to change both the world and how people live since it can automate procedures and learn from mistakes.

3. DEEP LEARNING TECHNIQUES

This section discusses the various deep neural network techniques. These strategies frequently use hierarchical structures with numerous levels of information processing to learn. Among the numerous hidden layers that are frequently observed in deep neural networks are the input and output layers. Reviewing the different training exercises available, that is (i) Supervised, an approach that is challenging and utilizes the use of tagged training data, and (ii) Unsupervised, an approach that examines unlabeled sets of data, is important before diving into the details of DL approaches.

3.1 SUPERVISED OR DISCRIMINATIVE LEARNING NETWORK

The term "supervised learning" refers to a process in which a supervisor doubles as an educator. The technique of instructing or training a computer system using labelled data is known as supervised learning. This suggests that the appropriate response has already been given to the given data. The machine is then given a new collection of examples so that the supervised learning algorithm may examine the training data (set of training examples) and provide an accurate output from labelled data. Discriminative deep architectures are frequently developed to give discriminative capability for pattern classification by modelling the posterior distributions of classes conditioned on observable data [29]. The three main categories of discriminative architectures are Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN or ConvNet), Recurrent Neural Networks (RNN), and their variations. Here, we'll briefly discuss these techniques.

3.1.1 MULTI-LAYER PERCEPTRON (MLP)

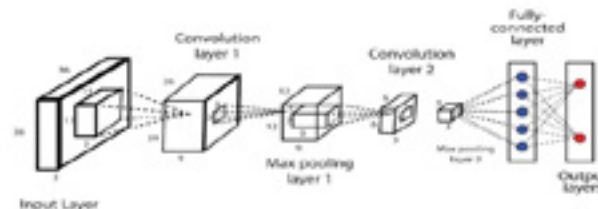
The feed-forward artificial neural network known as the Multi-layer perceptron (MLP) [30] is a technique for supervised instruction (ANN). The deep neural network (DNN) or deep learning base architecture are other names for it. Several activation functions, often referred to as transfer functions, including ReLU (Rectified Linear Unit), Tanh, Sigmoid, and Softmax determine an MLP network's output [32].

The most popular technique for training MLP is back-propagation [31], a supervised learning method that is frequently referred to as the fundamental component of a neural network. Throughout the training phase, a variety of optimization techniques are employed like Stochastic Gradient Descent (SGD), Limited Memory BFGS (L-BFGS), and Adaptive Moment Estimation (Adam).

3.1.2 CONVOLUTIONAL NEURAL NETWORK (CNN)

The commonly used deep learning architecture known as a convolutional neural network [33] was modelled after the visual brain of animals [34]. As seen in fig 2, Originally it had been used extensively for tasks involving object recognition, but it is currently also being investigated in areas such as object tracking [35], pose estimation [36], text detection and recognition [37], visual saliency detection [38], action recognition [39], scene labelling, and many more [40].

Since CNNs are particularly made to control the various 2D forms, they are frequently employed in visual identification, analyze clinical data, segmenting an image, processing language naturally, and many other applications [41]. Several CNN versions, including visual geometry group (VGG) [42], AlexNet [43], Xception [44], Inception [45], ResNet [46], etc., may be used in various application sectors depending on their learning capacity.



1. An illustration of Convolutional Neural Network.

3.1.3 RECURRENT NEURAL NETWORK (RNN)

Using sequential or time-series data, a different well-known neural network provides the result of one stage as input to the following step. The name for this neural network is recurrent neural network (RNN) [47]. Recurrent neural networks, like CNN and feed forward, learn from training input, but they stand out due to their "memory," which enables them to affect current input and output by consulting information from earlier inputs. While an RNN's output is reliant on what came before it in the sequence, a typical DNN assumes that inputs and outputs are independent of one another. However, because to the issue of declining gradients, standard networks with recurrence have difficulty in learning long data sequences. The popular recurrent network versions that the problems and perform effectively across a variety of real-life application areas are explored next.

3.1.4 LONG SHORT-TERM MEMORY (LSTM)

LSTMs are frequently used in video-based individual task re-ID and are capable of extracting temporal characteristics. Network for recurrent feature aggregation based on LSTM efficiently reduced interference brought on by background noise, shadowing, and recognition failure [48]. Among the first and shallowest LSTM nodes, it gathered cumulative discriminative characteristics. The temporal and geographical characteristics of the sections that include the probe pictures put were learned by the breakdown of a video sequence into multiple pieces [49]. The number of identical pedestrians in the sample is decreased by using this strategy, which also makes it simpler to identify similarity traits. Both of the aforementioned methods process each video frame independently. The duration of the video sequence typically has an impact on the characteristics that LSTM extracts. The RNN cannot catch the temporal signals of small details in the picture because it only creates temporal connections on high-level characteristics [50]. Therefore, research into a more effective technique for extracting spatial-temporal characteristics is still necessary.

3.1.5 GATED RECURRENT UNITS (GRUS)

A popular gating-based variation of the recurrent network techniques to monitor and regulate the flow of information between neural network units is the Gated Recurrent Unit (GRU) [51]. A reset gate and an update gate are all that the GRU has, as seen in Fig. 3, making it less complex than an LSTM. The primary difference between the two devices is the number of gates: an LSTM has three gates compared to a GRU's two (the reset and update gates). The GRU's characteristics allow dependencies from long data sequences to be collected adaptively without removing information from previous portions of the sequence. GRU is a little more compact approach as a consequence, often providing comparable results, and is significantly faster to compute [52].

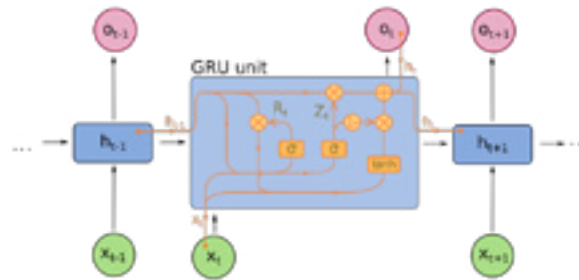


Figure 3. A Gated Recurrent Unit's Basic Structure with a Reset and Update Gate.

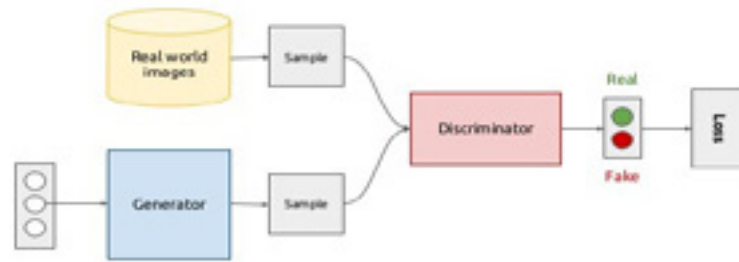
3.2 UNSUPERVISED LEARNING NETWORK

DL approaches are widely used to explain the combined statistical parameters of the available data and the classes that they belong to, as well as the higher predictive properties or features for pattern recognition or synthesis [53]. Since the methods under this category include frequently used to learn features or data generation and representation, they are fundamentally utilized for unsupervised learning [54]. Since generative modeling maintains the correctness of the discriminative model, it may be used as a preliminary step for supervised learning tasks as well. For generative learning or unsupervised learning, deep neural network algorithms including the Generative Adversarial Network (GAN), Autoencoder (AE), Restricted Boltzmann Machine (RBM), Self-Organizing Map (SOM), and Deep Belief Network (DBN), as well as its variations, are often utilized.

3.2.1 GENERATIVE ADVERSARIAL NETWORK (GAN)

.GANs make use of neural networks' capacity to train a function that can simulate a distribution as closely as feasible to the real thing. They are particularly capable of producing synthetic pictures with great visual fidelity and do not rely on prior assumptions about the distribution of the data. This important characteristic enables the application of GANs to any imbalance issue in computer vision tasks. GANs give a technique to alter the original picture in addition to being able to create a false image. There are several GANs with different strengths that have been published in the literature to address the imbalance issue in computer vision tasks. For example, a specific type of GANs called AttGAN [55], IcGAN [56], ResAttrGAN [57], etc., is frequently employed for tasks involving modifying face attributes.

GANs are comprised of two neural networks, as shown in Fig. 4. The discriminator D predicts the chance that a following sample will be taken from real data as opposed to data given by the generator G, which generates new data with features similar to the original data. The generator and discriminator in GAN modeling are then instructed to interact with one another. Healthcare, computer vision, data augmentation, video production, voice synthesis, epidemic prevention, traffic control, network security, and many more fields might all benefit from the utilization of GAN networks. In general, GANs have demonstrated to be a solid field of independent data expansion and a solution to problems requiring generative techniques.



2. Generative Adversarial Networks Framework [58].

3.2.2 AUTO-ENCODER (AE)

A well-known auto-encoder (AE) unsupervised learning approach that makes use of neural networks to learn representations [59]. Data reduction describes the depiction of a set of data. High-dimensional data are often processed using auto-encoders. Three parts make up an autoencoder: an encoder, a code, and a decoder. The encoder creates the code that the decoder uses to reproduce the input by compressing it. Furthermore, generative data models have been learned using the AEs [60]. Numerous unsupervised learning techniques, including dimension reduction, extraction of features, useful coding, dynamic modeling, noise removal, outlier or predictive modeling, etc., primarily rely on the auto-encoder. [59, 61]

3.2.3 KOHONEN MAP OR SELF-ORGANIZING MAP (SOM)

Another unsupervised learning method for constructing a low-dimensional (usually two-dimensional) representation of a higher-dimensional data set while preserving the topological structure of the data is the Self-Organizing Map (SOM) or Kohonen Map [62]. SOM is a neural network-based method for dimension reduction in clustering [63]. We can display huge datasets and identify likely clusters by using a SOM, which constantly moves a dataset's topological layout by bringing its neurons close to the data points inside the dataset. The input layer is the initial layer of a SOM, followed by the output layer, also known as the feature map, is the second layer. SOMs use competitive learning, which makes use of a neighboring function to preserve the topological properties of the input space, in contrast to other neural network models that use error-correction learning, such as Backpropagation with gradient descent [64]. Sequence identification, sickness or health diagnosis, fault diagnosis, and virus or parasite attack detection are just a few of the many activities that SOM is frequently employed for [65]. The main advantage of using a SOM is that it facilitates the discovery and recognition of patterns in high-dimensional data.

3.2.4 BOLTZMANN MACHINE WITH RESTRICTIONS (RBM)

A generative statistical neural network with the ability to learn a likelihood function over its inputs is the Restricted Boltzmann Machine (RBM) [66]. Each node in a Boltzmann machine can be either visible or hidden and is linked to every other node. By understanding how the system functions normally, we can better comprehend anomalies. In RBMs, a subset of Boltzmann machines, there is a limit on the number of linkages between the accessible and deep layers [67]. Due to this constraint, training techniques for Boltzmann machines in general can be more effective than those for Boltzmann machines, such as the gradient-based contrastive divergence algorithm [68]. Among the various uses of RBMs are data reduction, categorization, prediction, content-based- based filtering, pattern recognition, subject modelling, and many others.

3.2.5 DEEP BELIEF NETWORK (DBN)

A Deep Belief Network (DBN) [69] is a multiple-layer adaptive visual model composed of so many unsupervised networks, such as AEs or RBMs, layered one on top of the other and using the hidden layer of each model as the input for the layer below it or connected sequentially. As a result, there are two types of DBNs: AE-DBNs, also called as stacked AE, and RBM-DBNs, also known as stacked

RBM. The AE-DBN consists of autoencoders, whereas Boltzmann machines with constraints constitute RBM-DBN, as was already indicated. The final objective is to create a descriptive divergence-based quicker unsupervised training method for every sub-network [70]. The deep structure of DBN allows it to store a hierarchical representation of incoming data. Network architectures for unsupervised feed-forward are trained using unlabeled data as the basic tenet of DBN, and the networks are subsequently fine-tuned using marked input. DBN's potentially significant benefits over traditional shallow learning networks is its capacity to identify specific patterns, with strengthens logic and the ability to distinguish between true and wrong data [71].

Thus, the strategies for generative learning that were previously explored frequently allow us to use research method to build a new set of the data. Deep generative models of these kinds with supervised or discriminative learning methods may benefit from this preparation and to guarantee model correctness because improving classifier generalization through unsupervised representation learning.

4. CHALLENGES AND OPEN ISSUES

The fundamental difficulty with Re-ID is the variance in a person's appearance across multiple cameras. Re-ID is challenging to make self-operating for a variety of reasons. Re-ID networks usually consist of two essential parts: the capture of a unique individual description and the process of comparing two models to see if they match or don't match. The capacity to automatically recognize and track individuals in photos or videos is necessary in order to develop a distinctive person description. Numerous difficulties and problems are apparent, and they will guide future research in the area of person Re-ID.

4.1 RE-ID DEPENDING ON DEPTH

Depth photos capture the bones and contours of the body. Re-ID is made possible by this, which is crucial for applications involving individualized human contact in lighting and clothing variations [88]. In [72], a paradigm based on recurrent attention is put out to solve individual identification based on depth. Convolutional and recurrent neural networks are used to locate tiny, exclusionary localized parts of the body in a reinforcement learning framework.

4.2 RE-ID USING VISIBLE-INFRARED TECHNOLOGY

Visible-Infrared Re-ID handles the cross-modality matching between the noticeable and thermal pictures [88]. Because only infrared cameras can take photographs in low-light conditions, it is essential [73]. Along with the cross-modality shared embedding learning [74] also looks into the classifier level discrepancy. Recent methods [75] decrease cross-modality disparity at both the picture and feature level by creating cross-modality person photographs and applying the GAN approach. [76] models the cross-modal reconstruction using a hierarchy elements. [77] presents a dual-attentive aggregate learning strategy to identify multi-level links.

4.3 CROSS-RESOLUTION RE-ID

Taking into consideration the major resolution variations [78], Cross-Resolution Re-ID [88] compares images with different resolutions. The high-resolution human pictures are produced in a cascaded fashion using a cascaded SR-GAN [79], which also incorporates the identification data. The adversarial learning method is used by Li et al. [80] to create representations of pictures that are independent of resolution.

4.4 LABEL NOISE FOR RE-ID

It is typically hard to eliminate label noise when there is an annotation issue [88]. To prevent label overfitting problems, Zheng et al. use a label smoothing algorithm [81]. To effectively learn a Re-ID model while avoiding label noise and the consequences of data with high characteristic ambiguity should be mitigated, a Distribution Net (DNet) is described in [82] that encodes the feature uncertainty.

For each identity, there aren't enough data for powerful Re-ID model training, unlike the generic classification issue [83]. It is also more challenging to learn the potent Re-ID model because of the unidentified new identities.

4.5 MULTI-CAMERA DYNAMIC NETWORK

The constantly updated multi-camera network [84], which necessitates model change for new cameras or probes, is another challenging issue. The Re-ID model may be updated and the representation can be tailored for different probing galleries employing an adaptive learning method with humans in the loop[85]. Active learning was a component of early research on continuous Re-ID in multi-camera networks [86]. [87] Introduces an approach for flexibility to adapt relying on the selective use of limited, non-redundant samples. Utilizing the ideal source camera theory and a geodesic flow kernel serve as the foundation for the development of a transitive inference strategy. An open-world person Re-ID system adds a number of contextual limitations (such Camera Topology) while dealing with large crowds and social interactions[88].

4.6 FEATURE LEARNING

High-level semantic representations of a person's traits, such their hair, gender, and age, can resist several environmental changes. For deep learning-based person Re-ID systems like those in [89], some research has used these traits to fill the gap between the photos and high-level conceptual data. Since it delivered on its predictions, feature recognition is one of the next possibilities.

4.7 ARCHITECTURE FOR AUTOMATED RE-ID

An advanced learning model's architectures must be manually created, which takes time, effort, and is prone to mistakes. The approach of automating architectural engineering, known as neural architecture search (NAS) [90], has recently been applied to address this issue. The study of NAS is currently receiving greater attention. Therefore, one of the essential aspects that must be considered in future research is the use of NAS for person Re-ID activities, as the majority of NAS techniques don't assure that there commended CNN is suitable for person Re-ID tasks.

4.8 ACCURACY VERSUS EFFICIENCY

Large models are typically employed to obtain the greatest accuracy, but they can be time and memory intensive, which reduces their usefulness, particularly when used to mimic real-time video monitoring systems. The majority of modern models did not consider CPU speed and memory capacity into account in order to increase accuracy. Authors who work in these fields must strike a compromise between processing speed and ranking accuracy.

4.9 LIGHTWEIGHT MODEL

The creation of a lightweight Re-ID model is another approach for dealing with the scalability problem. The issue of changing the network topology to create a light model is investigated [91, 92]. Another strategy is to employ model distillation. A system for multi-teacher customizable comparison reduction, for instance, is provided in [93], which in the absence of a primary data source, trains a user-specified lightweight student model from a number of teacher models.

5. DATASETS AND EVALUATION

Individuals' appearances vary greatly depending on Lighting, stances, view angles, scales, and camera resolutions may all vary while using different cameras. Visual ambiguities are further increased by elements like occlusions, a crowded background, and articulated figures. Therefore, it is crucial to gather data that successfully captures these aspects in order to create viable Re-ID approaches. In addition to good quality data that replicates actual circumstances, it is essential to compare and assess the Re-ID methodologies that are developed and find ways to enhance the methodology and databases.

5.1 DATASETS BASED ON IMAGES

Person re-ID using images, there have been a variety of datasets; the most common datasets are listed below.

VIPeR[94]: It is made up of 1,264 photos for 632 people and was taken by 2 non-overlapping cameras.

CUHK01 [95]: It is made up of 3,884 photographs for 971 individuals that were recorded by two separate cameras on a university campus.

Market-1501 [96]: It is comprised of 32,643 photos for each of the 1,501 people that make up the sample, which was taken from the front of a shop using 2 to 6 separate cameras.

DukeMTMC-ReID[97]: The dataset consists of 46,261 photos for 1852 individuals that were captured by 8 non-overlapping cameras on the Duke University campus.

Kinect-REID [98]: It has 71 person sequences that were recorded at the authors' department.

RGBD-ID [99]: There are four groups with various viewpoints, each with the same 80 people. It is produced on several days and has various aesthetic variants.

RegDB[100]: 412 people are represented by 4120 RGB photos and 4120 thermal photographs. Two types of cameras are used to capture them.

SYSU-MM01 [101]: It comprises of 491 people's 15,792 infrared photographs and 287,628 RGB photos. Six cameras, including two infrared cameras and four RGB cameras, are used to capture them from the writers' section.

5.2 DATASETS BASED ON VIDEOS

PRID2011 [102]: It is constituted of 24541 photos for 934 individuals from 600 recordings taken from two separate cameras in an airport's multi-camera network.

iLIDSVID[103]: It is made up of 600 movies shot by 2 non-overlapping airport cameras and 42495 photos for 300 people.

MARS [104]: The greatest video-based individual Re-ID dataset resides in this one. It is made up of around 1191003 photos for 1261 individuals from 200 recordings that were captured by 2 to 6 non-overlapping cameras.

RPIfield [105]: It consists of 601,581 images for 112 individuals, captured by 2 separate cameras on an open field at a college.

5.3 EVALUATION METRICS

Cumulative matching characteristics (CMC) [106] and mean average precision (mAP) are the two mostly used metrics for assessing Re-ID systems [107].

6. CONCLUSION

We have discussed the subject of human re-identification, as well as difficult problems and a summary of recent research in the discipline of person recognition, in this paper. Both closed set and open set Re-ID tasks have been taken into consideration. The approaches employed have been grouped, and their advantages and disadvantages have been covered. Additionally, we have outlined the benefits and drawbacks of the various Re-ID datasets. Popular Re-ID assessment methods are briefly discussed, along with potential expansions. In conclusion, person Re-ID is a broad and difficult field with much of space for growth and research. An effort is made in this work to give a concise overview of the Re-ID problem, its limitations, and related problems.

REFERENCES

- (1) Y.-C. Chen, X. Zhu, W.-S. Zheng, and J.-H. Lai, "Person reidentification by camera correlation aware feature augmentation," *IEEE TPAMI*, vol. 40, no. 2, 2018.
- (2) N. Gheissari, T. B. Sebastian, and R. Hartley, "Person reidentification using spatiotemporal appearance," in *CVPR*, 2006, pp. 1528–1535.
- (3) J. Almazan, B. Gajic, N. Murray, and D. Larlus, "Re-id done right: towards good practices for person re-identification," *arXiv preprint arXiv:1801.05339*, 2018.
- (4) T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *ECCV*, 2014.
- (5) M. Ye, C. Liang, Z. Wang, Q. Leng, J. Chen, and J. Liu, "Specific person retrieval via incomplete text description," in *ACM ICMR*, 2015, pp. 547–550.
- (6) S. Karanam, Y. Li, and R. J. Radke, "Person re-identification with discriminatively trained viewpoint invariant dictionaries," in *ICCV*, 2015, pp. 4516–4524.
- (7) X. Li, W.-S. Zheng, X. Wang, T. Xiang, and S. Gong, "Multi-scale learning for low-resolution person re-identification," in *ICCV*, 2015, pp. 3765–3773.
- (8) Y. Huang, Z.-J. Zha, X. Fu, and W. Zhang, "Illumination-invariant person re-identification," in *ACM MM*, 2019.
- (9) Y.-J. Cho and K.-J. Yoon, "Improving person re-identification via pose-aware multi-shot matching," in *CVPR*, 2016, pp. 1354–1362.
- (10) H. Huang, D. Li, Z. Zhang, X. Chen, and K. Huang, "Adversarially occluded samples for person re-identification," in *CVPR*, 2018, pp. 5098–5107.
- (11) A. Wu, W.-s. Zheng, H.-X. Yu, S. Gong, and J. Lai, "Rgb-infrared cross-modality person re-identification," in *ICCV*, 2017.
- (12) C. Song, Y. Huang, W. Ouyang, and L. Wang, "Mask-guided contrastive attention model for person re-identification," in *CVPR*, 2018, pp. 1179–1188.
- (13) A. Das, R. Panda, and A. K. Roy-Chowdhury, "Continuous adaptation of multi-camera person identification models through sparse non-redundant representative selection," *CVIU*, vol. 156, pp. 66–78, 2017.
- (14) J. Garcia, N. Martinel, A. Gardel, I. Bravo, G. L. Foresti, and C. Micheloni, "Discriminant context information analysis for post-ranking person re-identification," *IEEE Transactions on ImageProcessing*, vol. 26, no. 4, pp. 1650–1665, 2017.
- (15) W.-S. Zheng, S. Gong, and T. Xiang, "Towards open-world person re-identification by one-shot group-based verification," *IEEE TPAMI*, vol. 38, no. 3, 2015.
- (16) A. Das, R. Panda, and A. Roy-Chowdhury, "Active image pair selection for continuous person re-identification," in *ICIP*, 2015, pp. 4263–4267.
- (17) J. Song, Y. Yang, Y.-Z. Song, T. Xiang, and T. M. Hospedales, "Generalizable person re-identification by domain-invariant mapping network," in *CVPR*, 2019.
- (18) A. Das, A. Chakraborty, and A. K. Roy-Chowdhury, "Consistent re-identification in a camera network," in *ECCV*, 2014, pp. 330–345.
- (19) Q. Yang, A. Wu, and W. Zheng, "Person re-identification by contour sketch under moderate clothing change," *IEEE TPAMI*, 2019.

- (20) J. Sivic, C.L. Zitnick, R. Szeliski, Finding people in repeated shots of the same scene, Proceedings of the British Machine Vision Conference, 2006, pp. 909–918.
- (21) M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In 2010 IEEE Computer Society Conference
- (22) S. Liao and S. Z. Li. Efficient psd constrained asymmetric metric learning for person re-identification. In 2015 IEEE International Conference on Computer Vision (ICCV), pages 3685–3693, 2015.
- (23) Liang Zheng, Yi Yang, and Alexander G. Hauptmann. Person re-identification: Past, present and future. ArXiv, abs/1610.02984, 2016.
- (24) Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In AAAI, 2019.
- (25) Kaiming He, Xiangyu Zhang, ShaoqingRen, and Jian Sun. Deep residual learning for image recognition, 2015.
- (26) J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei- Fei. Imagenet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255, 2009.
- (27) Sarker IH. Data science and analytics: an overview from datadrivensmart computing, decision-making and applications perspective. SN Comput Sci. 2021.
- (28) Sarker IH. Machine learning: Algorithms,real-world applications and research directions. SN Computer. Science. 2021;2(3):1–21.
- (29) Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Trans Signal Inf Process. 2014; p. 3.
- (30) Pedregosa F, Varoquaux G, Gramfort A, Michel V, ThirionB,Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. Scikit-learn: machine learning in python. J Mach Learn Res.2011;12:2825–30.
- (31) Han J, Pei J, Kamber M. Data mining: concepts and techniques. Amsterdam: Elsevier; 2011.
- (32) Sarker IH. Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. SN Computer.Science. 2021;2(3):1–16.
- (33) Ramachandran R, Rajeev DC, Krishnan SG, P Subathra, Deep learning an overview, IJAER, Volume 10, Issue 10, 2015, Pages 25433-25448.
- (34) D. H. Hubel and T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, The Journal of physiology, 1968.
- (35) J. Fan, W. Xu, Y. Wu, and Y. Gong, Human tracking using convolutional neural networks, Neural Networks, IEEE Transactions, 2010.
- (36) A. Toshev and C. Szegedy, Deep -pose: Human pose estimation via deep neural networks, in CVPR, 2014.
- (37) M. Jaderberg, A. Vedaldi, and A. Zisserman, Deep features for text spotting, in ECCV, 2014.
- (38) R. Zhao, W. Ouyang, H. Li, and X. Wang, Saliency detection by multicontext deep learning, in CVPR, 2015.
- (39) J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, Decaf: A deep convolutional activation feature for generic, 2014

- (40) Nithin, D Kanishka and Sivakumar, P Bagavathi, *Generic Feature Learning in Computer Vision*, Elsevier, Vol.58, Pages202-209, 2015.
- (41) LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86(11):2278–324.
- (42) He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell*. 2015;37(9):1904–16.
- (43) Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. 2012.
- (44) Chollet F. Xception: Deep learning with depthwise separable convolutions. In: *Proceedings of the IEEE Conference on computer vision and pattern recognition*, 2017
- (45) [45] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on computer vision and pattern recognition*, 2016
- (46) [46] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *Proceedings of the IEEE Conference on computer vision and pattern recognition*, 2015.
- (47) Dupond S. A thorough review on the current advance of neural network structures. *Annu Rev Control*. 2019.
- (48) Y. Yan, B. Ni, Z. Song, C. Ma, Y. Yan, X. Yang, Person reidentification via recurrent feature aggregation, *Proceedings of the European Conference on Computer Vision*, 2016.
- (49) D. Chen, H. Li, T. Xiao, S. Yi, X. Wang, Video person reidentification with competitive snippet-similarity aggregation and co-attentive snippet embedding, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1169– 1178
- (50) J. Li, S. Zhang, T. Huang, Multi-scale 3d convolution network for video based person re-identification, *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 2019, pp. 8618– 8625
- (51) Chung J, Gulcehre C, Cho KH, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.
- (52) Gruber N, Jockisch A. Are gru cells more specific and lstm cells more sensitive in motive classification of text? *Front ArtifIntell*. 2020;3:40.
- (53) Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Trans Signal Inf Process*. 2014.
- (54) Da'u A, Salim N. Recommendation system based on deep learning methods: a systematic review and new directions. *Artif Intel Rev*. 2020;53(4):2709–48.
- (55) He Z, Zuo W, Kan M, Shan S, Chen X. AttGAN: Facial attribute editing by only changing what you want. *IEEE transactions on image processing* . 2019;28:5464–78.
- (56) Perarnau G, van de Weijer J, Raducanu B, Álvarez JM. Invertible Conditional GANs for image editing. *Conference on Neural Information Processing Systems* . 2016.
- (57) Tao R, Li Z, Tao R, Li B. ResAttr-GAN: Unpaired deep residual attributes learning for multi-domain face image translation. *IEEE Access* . 2019;7:132594–608.
- (58) Moacir A. Ponti, Leonardo S. F. Ribeiro, Tiago S. Nazare, Tu Bui, John Collomosse “Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask” <https://www.researchgate.net/publication/322413149>

- (59) Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. Cambridge: MIT Press; 2016.
- (60) Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. *Neurocomputing*. 2017;234:11–26.
- (61) Zhang G, Liu Y, Jin X. A survey of autoencoder-based recommender systems. *Front Comput Sci*. 2020;14(2).
- (62) Kohonen T. The self-organizing map. *Proc IEEE*. 1990;78(9):1464–80.
- (63) [63] Sarker IH, Salah K. Appspred: predicting context-aware smartphone apps using random forest learning. *Internet of Things*. 2019;8:100106.
- (64) Han J, Pei J, Kamber M. *Data mining: concepts and techniques*. Amsterdam: Elsevier; 2011.
- (65) Kohonen T. Essentials of the self-organizing map. *Neural Netw*. 2013;37:52–65.
- (66) [66] Marlin B, Swersky K, Chen B, Freitas N. Inductive principles for restricted boltzmann machine learning. In: *Proceedings of the Thirteenth International Conference on artificial intelligence and statistics*, p. 509–16. *JMLR Workshop and Conference Proceedings*, 2010.
- (67) Memisevic R, Hinton GE. Learning to represent spatial transformations with factored higher-order boltzmann machines. *Neural Comput*. 2010;22(6):1473–92.
- (68) Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18(7):1527–54.
- (69) Hinton GE. Deep belief networks. *Scholarpedia*. 2009;4(5):5947.
- (70) Hinton GE, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets. *Neural Comput*. 2006;18(7):1527–54.
- (71) Ren J, Green M, Huang X. From traditional to deep learning: fault diagnosis for autonomous vehicles. In: *Learning control*. Elsevier. 2021; p. 205–19.
- (72) A. Haque, A. Alahi, and L. Fei-Fei, “Recurrent attention models for depth-based person identification,” in *CVPR*, 2016, pp. 1229–1238.
- (73) M. Ye, Z. Wang, X. Lan, and P. C. Yuen, “Visible thermal person re-identification via dual-constrained top-ranking,” in *IJCAI*, 2018, pp. 1092–1099.
- (74) M. Ye, J. Shen, and L. Shao, “Visible-infrared person re-identification via homogeneous augmented tri-modal learning,” *IEEE TIFS*, 2020.
- (75) Z. Wang, Z. Wang, Y. Zheng, Y.-Y. Chuang, and S. Satoh, “Learning to reduce dual-level discrepancy for infrared-visible person re-identification,” in *CVPR*, 2019, pp. 618–626.
- (76) S. Choi, S. Lee, Y. Kim, T. Kim, and C. Kim, “Hi-cmd: Hierarchical cross-modality disentanglement for visible-infrared person re-identification,” in *CVPR*, 2020, pp. 257–266.
- (77) M. Ye, J. Shen, D. J. Crandall, L. Shao, and J. Luo, “Dynamic dual-attentive aggregation learning for visible-infrared person re-identification,” in *ECCV*, 2020.
- (78) X. Li, W.-S. Zheng, X. Wang, T. Xiang, and S. Gong, “Multi-scale learning for low-resolution person re-identification,” in *ICCV*, 2015, pp. 3765–3773.
- (79) Z. Wang, M. Ye, F. Yang, X. Bai, and S. Satoh, “Cascaded sr-gan for scale-adaptive low resolution person re-identification.” in *IJCAI*, 2018, pp. 3891–3897.

- (80) Y.-J. Li, Y.-C. Chen, Y.-Y. Lin, X. Du, and Y.-C. F. Wang, "Recover and identify: A generative dual model for cross-resolution person re-identification," in ICCV, 2019, pp. 8090–8099.
- (81) Z. Zheng, L. Zheng, and Y. Yang, "Unlabeled samples generated by gan improve the person re-identification baseline in vitro," in ICCV, 2017, pp. 3754–3762.
- (82) T. Yu, D. Li, Y. Yang, T. Hospedales, and T. Xiang, "Robust person re-identification by modelling feature uncertainty," in ICCV, 2019, pp. 552–561.
- (83) M. Ye and P. C. Yuen, "Purifynet: A robust person reidentification model with noisy labels," IEEE TIFS, 2020.
- (84) A. Das, A. Chakraborty, and A. K. Roy-Chowdhury, "Consistent re-identification in a camera network," in ECCV, 2014.
- (85) N. Martinel, A. Das, C. Micheloni, and A. K. Roy-Chowdhury, "Temporal model adaptation for person re-identification," in ECCV, 2016.
- (86) A. Das, R. Panda, and A. Roy-Chowdhury, "Active image pair selection for continuous person re-identification," in ICIP, 2015.
- (87) A. Das, R. Panda, and A. K. Roy-Chowdhury, "Continuous adaptation of multi-camera person identification models through sparse non-redundant representative selection," CVIU, vol. 15.2017.
- (88) Mang Ye, JianbingShen, Gaojie Lin, Tao Xiang, Ling Shao, Steven C. H. Hoi, Deep Learning for Person Re-identification: A Survey and Outlook, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE-2021
- (89) C. Su, S. Zhang, J. Xing, W. Gao, and Q. Tian, "Multi-type attributes driven multi-camera person re-identification," Pattern Recognit., vol. 75, pp. 77_89, Mar. 2018.
- (90) T. Elsken, J. H. Metzen, and F. Hutter, "Correction to: Neural architecture search," in Automated Machine Learning: Methods, Systems, Challenges, F. Hutter, L. Kotthoff, and J. Vanschoren, Eds. Cham, Switzerland: Springer, 2019.
- (91) W. Li, X. Zhu, and S. Gong, "Harmonious attention network for person re-identification," in CVPR, 2018, pp. 2285–2294.
- (92) K. Zhou, Y. Yang, A. Cavallaro, and T. Xiang, "Omni-scale feature learning for person re-identification," in ICCV, 2019, pp. 3702–3712.
- (93) A. Wu, W.-S. Zheng, X. Guo, and J.-H. Lai, "Distilled person reidentification: Towards a more scalable system," in CVPR, 2019, pp. 1187–1196.
- (94) D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in Proc. 10th Int. Workshop Perform. Eval. Tracking Surveill. (PETS), vol. 3, 2007, pp. 41_47.
- (95) W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," in Computer Vision ACCV (Lecture Notes in Computer Science), vol. 7724. Berlin, Germany: Springer, 2013, pp. 31_44.
- (96) L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1116_1124.
- (97) E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in Computer Vision ECCV 2016 Workshops (Lecture Notes in Computer Science), vol. 9914. Cham, Switzerland: Springer, 2016, pp. 17_35.

- (98) F. Pala, R. Satta, G. Fumera, and F. Roli, "Multimodal person re-identification using RGB-D cameras," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 4, pp. 788_799, Apr. 2016.
- (99) I. B. Barbosa, M. Cristani, A. Del Bue, L. Bazzani, and V. Murino, "Re-identification with RGB-D sensors," in *Computer Vision_ECCV2012. Workshops and Demonstrations*, A. Fusiello, V. Murino, and R. Cucchiara, Eds. Berlin, Germany: Springer, 2012, pp. 433_442.
- (100) D. T. Nguyen, H. G. Hong, K. W. Kim, and K. R. Park, "Person recognition system based on a combination of body images from visible light and thermal cameras," *Sensors*, vol. 17, no. 3, p. 605, 2017.
- (101) A. Wu, W.-S. Zheng, H.-X. Yu, S. Gong, and J. Lai, "RGB-infrared crossmodality person re-identification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017.
- (102) M. Hirzer, C. Belezni, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," in *Image Analysis (Lecture Notes in Computer Science)*, vol. 6688. Berlin, Germany: Springer, 2011, pp. 91_102.
- (103) T. Wang, S. Gong, X. Zhu, and S. Wang, "Person re-identification by video ranking," in *Computer Vision_ECCV*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham, Switzerland: Springer, 2014, pp. 688_703.
- (104) L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "MARS: A video benchmark for large-scale person re-identification," in *Computer Vision_ECCV (Lecture Notes in Computer Science)*, vol. 9910. Cham, Switzerland: Springer, 2016, pp. 868_884.
- (105) M. Zheng, S. Karanam, and R. J. Radke, "RPI_eld: A new dataset for temporally evaluating person re-identification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018.
- (106) X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *ICCV*, 2007.
- (107) L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *ICCV*, 2015, pp. 1116_1124.