PERSON RE-IDENTIFICATION USING CONVOLUTIONAL NEURAL NETWORK

Dr.G.Sivakumar[1], S. Bharanidharan[2], K. V. Dharun Kumar[3], R. Gokul[4]

Department of CSE, Erode Sengunthar Engineering College,

Erode, Tamil Nadu, India.

ABSTRACT

Machine learning, which is basically a three-layer neural network, includes deep learning as a subset. While they fall far short of the human brain's capacity to "learn" from vast quantities of data, these neural networks make an effort to mimic its behavior. An intelligent image surveillance technique called individual Re-Identification (ReID) may identify the same individual across several cameras. Occlusion, shifting camera angles, and human stance variations make this work quite difficult. Unconstrained spatial misalignment between picture pairs resulting from changes in view angle and pedestrian position, together with label noise resulting from clustering, are significant challenges in person-recognition identification (ReID). To solve this issue, Convolutional Neural Network (CNN) is a preprocessing method based on reinforcement learning that learns task-specific sequential spatial correspondences for various image pairs through local pairwise internal representation interactions. It is the suggested method for person ReID and is carried out based on the best features. Next, provide some instances of frequently used datasets, evaluate the benefits and drawbacks of different approaches, and compare how well particular algorithms work on newly collected picture datasets. Deep learning models for face recognition may then be trained using the newly created photos from CNN. CNNs are very helpful for image identification, image classification, and computer vision (CV) applications because they provide incredibly accurate results, especially when a large amount of data is involved. As the object data passes through the CNN's several layers, the CNN also picks up the item's characteristics via repeated repeats. The suggested approach provides accuracy on 96.0% and 89.0%, respectively, in comparison to the current technique.

Keywords: PERSON RE-IDENTIFICATION, DEEP METRIC LEARNING, LOCAL FEATURE LEARNING, GENERATIVE ADVERSARIAL LEARNING, SEQUENCE FEATURE LEARNING

1. INTRODUCTION

Person re-identification (re-ID) is a difficult computer vision job where the goal is to identify people across several photos or video frames taken by discontinuous or non-overlapping cameras. Its potential uses in public safety, person monitoring, and video surveillance have drawn a lot of interest in recent years. Person re-ID seeks to efficiently and reliably match a person's query picture with the matching photographs throughout a vast gallery collection, even in the face of notable alterations in the person's look, position, lighting, and occlusion. Person re-identification techniques have traditionally concentrated on creating feature representations that are both discriminative and invariant to different appearance-affecting circumstances. In order to match and rank people, these methods usually extract low-level visual data like color, texture, and form and train discriminative models. But even with these great advances, human re-identification is still a tough task, mainly because of the inherent challenges in managing large-scale fluctuations and collecting fine-grained features in real-world settings. Deep learning methods for human re-identification have garnered increasing attention in the last several years. Prominent achievements in computer vision tasks such as object identification and picture classification have been achieved using deep neural networks. Researchers have surpassed the capability of conventional approaches to attain state-of-the-art performance in person re-ID by using deep learning architectures.

1.1 PERSON RE-IDENTIFICATION

Person re-identification (re-ID) is a computer vision challenge in which subjects are identified and matched across several pictures or video frames taken by discontinuous or non-overlapping cameras. Its applications in a variety of fields, including as public safety, person monitoring, and video surveillance, have drawn a lot of interest in recent years. Person re-ID aims to precisely and quickly identify a person from a query picture by matching it to related photographs in a gallery collection, even when there are notable differences in look, posture, lighting, and occlusion. Comparing person re-identification to other computer vision tasks, there are distinct problems involved. The work entails managing large-scale alterations, such as changes in clothes, camera perspectives, and ambient circumstances, as well as gathering minute characteristics about each subject. Furthermore, real-world settings provide additional challenges to the work as the gallery collection may include a high number of photos with incorrect labeling or overlapping identities. Person re-ID techniques historically matched and ranked people using manually created feature representations, such as color histograms, texture descriptors, or geometric characteristics. These methods, however, often failed to extract the nuanced and discriminative information required for precise individual re-identification.

1.2 DEEP METRIC LEARNING

Learning representations or embeddings that capture the similarity or dissimilarity between data samples is the main goal of the deep metric learning area of deep learning. Deep metric learning tries to generate a feature space where similar samples are closer to each other and dissimilar samples are further away, in contrast to standard deep learning tasks like object recognition or image classification, where the objective is to find objects in an image or give a label. Many applications, including as image retrieval, face recognition, person re-identification, and similarity-based recommendation systems, need deep metric learning. A reliable and discriminative data representation that can precisely gauge the similarity between several samples is crucial for these kinds of activities. Deep learning approaches are used because traditional handmade features often fail to capture the intricate connections and variations contained in the data. Deep metric learning uses deep neural network designs to learn highly discriminative and expressive feature embedding, such as triplet networks, Siamese networks, and convolutional neural networks (CNNs). The goal of training these networks using pairs or triplets of examples is to increase the distance between different samples in the embedding space and decrease the distance between similar samples.

1.3 LOCAL FEATURE LEARNING

A key component of computer vision is local feature learning, which is the process of identifying and encoding discriminative information from certain areas or patches of an image. Local feature learning seeks to identify local patterns, structures, and textures that might be important for a variety of visual tasks, including object detection, picture matching, and image retrieval. This is in contrast to global feature representations, which take the whole image into account. Usually, local features—also called key points or interest points—are retrieved from tiny picture areas. These focal points are areas inside a picture, such corners, edges, or blobs, that show clear, recurring visual patterns. Local feature learning techniques may provide resilience to changes in size, rotation, lighting, and other variables often seen in real-world pictures by concentrating on these relevant areas. In the past, local information was often extracted from pictures using manually created local feature descriptors like Scale-Invariant Feature Transform (SIFT), Speeded-Up Robust Features (SURF), or Local Binary Patterns (LBP). These descriptors allow for quick and accurate matching and identification by encoding the look and spatial connections of pixels within the immediate area. Direct learning of local feature representations from data has been more popular with the introduction of deep learning. By using their capacity to develop hierarchical representations, deep learning-based techniques like convolutional neural networks (CNNs) have shown impressive performance in collecting rich and discriminative local data. Large-scale datasets are used to train these networks so they can automatically learn feature representations that are remarkably resistant to changes and modifications.

1.4 GENERATIVE ADVERSARIAL LEARNING

The subject of generative modeling has undergone a revolution because to the potent machine learning framework known as generative adversarial learning, or GAN. It entails training a generator and a discriminator neural network in a competitive environment. The discriminator's goal is to discern between created and genuine samples, while the generator's is to generate realistic examples. Since its introduction by Ian Goodfellow and colleagues in 2014, the notion of GANs has drawn a lot of interest because of its capacity to produce very varied and realistic synthetic data in a variety of domains, including text, pictures, and even music. The generator network in a GAN creates artificial data samples that resemble the distribution of actual data by using random noise as input. In contrast, the discriminator network is taught to identify whether a particular sample is produced or genuine. Iterative training is used to train the discriminator and generator. The discriminator gains proficiency in differentiating between produced and genuine samples, while the generator tries to make ever more realistic examples that may trick the discriminator.

1.5 SEQUENCE FEATURE LEARNING

Within machine learning and deep learning, the area of sequence feature learning is concerned with extracting meaningful representations from sequential data. Sequential data includes time series, audio signals, spoken language, and genetic sequences—any data where the temporal relationships and order of items are significant. Knowing how to interpret and extract information from sequences is crucial for many real-world applications. For instance, accurate transcription in voice recognition depends on the audio signals' sequential character. For tasks like language modeling and machine translation in natural language processing, it is essential to capture the sequential relationships between words. Analyzing genetic sequences in bioinformatics enables the discovery of patterns linked to gene expression and illness diagnosis. Conventional methods of sequence analysis often depended on manually created features or domain-specific expertise.

2. LITERATURE SURVEY

Wang Xiaogang [1], et al. As shown in this study, the interdisciplinary topic of intelligent multi-camera video surveillance involves computer vision, pattern recognition, signal processing, embedded computing, networking, and image sensors. This study examines the latest advancements in pertinent technologies from the standpoints of pattern recognition and computer vision. Multi-camera calibration, calculating camera network topology, multi-camera tracking, object re-identification, multi-camera activity analysis, and cooperative video surveillance using both active and static cameras are among the subjects addressed. They give thorough explanations of the technical difficulties they face as well as a comparison of various remedies. It focuses on how distinct modules link and work together in a variety of settings and application situations. Recent research indicates that some issues may be cooperatively resolved to increase accuracy and efficiency. The sizes and complexity of camera networks are growing, along with the congested and complex monitored settings, due to the rapid development of surveillance technologies. The topic of this article is how to deal with these new issues.One of the computer vision fields with the most active development has been intelligent video surveillance. The objective is to automatically identify, monitor, and identify items of interest, as well as comprehend and analyze their movements, from the vast quantity of films that security cameras have recorded. Numerous public and private settings may benefit from video surveillance, including homeland security, crime prevention, traffic management, accident prediction and detection, and home monitoring of sick, the elderly, and children. For these applications, it is necessary to monitor scenes from roads, train stations, parking lots, shops, shopping centers, and workplaces, both inside and outside.

Al Masada [2] et al. has suggested in this documentRecently, computer vision researchers have been paying greater attention to person re-identification systems, or person Re-ID. They have several uses, including those related to public safety, and are essential to intelligent visual surveillance systems. Person Re-ID systems are able to determine if an individual has been seen in an unrestricted setting by a non-overlapping camera on a sizable camera network. It is a difficult problem since a person looks differently depending on the camera angle and encounters several difficulties such lighting variations, occlusion, and position variation. In order to address the person Re-ID issue, several techniques for creating handmade features

have been created. Since deep learning has shown notable achievements in computer vision problems, several research have begun to employ deep learning techniques to improve the person's Re-ID performance in recent years. As a result, this study offers an overview of current research that suggests using deep learning to enhance person Re-ID systems. There is discussion of the public datasets that are used to assess these systems. In order to improve person Re-ID systems, the study concludes by discussing existing challenges and future directions that need to be taken into account.In the area of computer vision, Person Re-Identification, or Person Re-ID, has lately garnered scholarly interest. Strong intelligent video surveillance is in greater demand as a result of its significance for security objectives in contemporary society, including the prevention of crimes and terrorist acts, forensic examination, etc. Governments work very hard to advance surveillance technologies for public safety. One of the most crucial and significant responsibilities in intelligent video surveillance systems is automated monitoring and analysis of captured footage. But it takes time and effort for a person to do this monitoring. One important duty in intelligent video surveillance systems is person re-identification. It is described as the procedure used in multi-camera surveillance systems across different geographic locations to identify and recognize the same individual across a group of non-overlapping cameras.Since the movies are taken by non-overlapping cameras in various situations, Person Re-ID is a difficult problem. Therefore, it is not helpful to utilize main biometric data for this purpose, such as face. Studies concentrate on a person's look, but there is also a great deal of visual ambiguity brought on by intra- and inter-class differences. Person Re-ID is still a difficult process, but it's one of the crucial and significant tasks in intelligent video surveillance systems. Deep learning person Re-ID systems were covered in this survey. A generic architecture for both conventional and deep learning systems was presented. A lot of recent research has shifted to deep learning in order to get beyond the drawbacks of manual techniques.

Zhenget Liang [3] et al. Person re-identification, or re-ID, has been suggested in this system and has gained popularity in the community because of its research value and practical applications. It tries to identify an interesting subject in other cameras. The majority of reports in the early stages focused on small-scale assessment and hand-crafted algorithms. Deep learning algorithms that use vast amounts of data have been developed in recent years, along with large-scale datasets. We divide the majority of existing re-ID techniques into two groups, namely image-based and video-based, taking into account various tasks; for each task, both hand-crafted and deep learning systems will be examined. Additionally, two novel re-identification tasks—end-to-end re-ID and quick re-ID in very large galleries—that are far more applicable to real-world scenarios are explained and explored. This paper describes the key future directions in end-to-end re-ID and fast retrieval in large galleries, briefly touches on some significant but unexplored issues, and introduces the history of person re-ID and its relationship with image classification and instance retrieval. It also surveys a wide range of hand-crafted systems and large-scale methods in both image- and video-based re-ID. As said by Homer, Menelaus intended to appease the gods and make a safe return home, but he was stuck on his way home after the Trojan War. It was instructed to him to apprehend Proteus and compel him to divulge the solution. When Proteus emerged from the sea to sleep among the seals, Menelaus managed to capture him, despite Proteus changing into a lion, a snake, a leopard, water, and a tree. At last, Proteus was forced to give him an honest response. This may be one of the first tales of someone regaining their identity despite significant physical alteration. Alvin Planting gave one of the first definitions of re-identification in 1961 when speaking on the connection between mental states and behavior. He said, "To re-identify a particular, then, is to identify it as (numerically) the same particular as one encountered on a previous occasion." Thus, person re-identification has been investigated in a number of fields, including logic, psychology, and metaphysics, in study and documentation. Leibniz's Law, which states that "there cannot be separate objects or entities that have all their properties in common," serves as the foundation for all of these works. The job of person re-ID in the contemporary computer vision field has similarities to earlier research.

Redmonet al., Joseph.[4] has suggested using this methodWe are introducing some YOLO upgrades! To improve it, we made a number of little design adjustments. We also trained this new, very good network. Though somewhat larger than before, it is more precise. Still, don't worry, it's quick. You know, sometimes you just sort of phone it in for a

whole year? This year, I didn't do a lot of research. logged in to Twitter many times. played around a little with GANs. I was able to enhance YOLO because I carried over some of my momentum from the previous year. However, there's really nothing very fascinating—just a number of little adjustments to make it better. I have provided some little assistance to others with their study. That's really the reason we're here today. We need to reference a few of my haphazard edits to YOLO before we miss our camera-ready deadline, but we don't have a source. Prepare accordingly for a TECH REPORT! The best thing about tech reports is that everyone knows why they are here, so no introduction is necessary. Thus, this introduction's conclusion will serve as a guide for the remainder of the work. We'll start by explaining the situation with YOLOv3. After that, we'll inform you how we perform. We will also share with you some of the unsuccessful attempts we made. Lastly, we'll consider the significance of everything. Additionally, we use concatenation to combine our upsampled features with an earlier feature map from the network. By using this technique, we may extract finer-grained information from the previous feature map and more significant semantic information from the up-sampled features. In order to analyze this merged feature map, we then add a few additional convolutional layers. Eventually, we predict a tensor that is comparable but now twice as large. To anticipate boxes for the final scale, we run the identical design through one more time. As a result, all of the earlier calculation and the fine-grained data from the network's early stages are beneficial to our forecasts for the third scale.Each box uses multilabel classification to anticipate which classes the enclosing box could include. We find that a softmax is not needed for decent performance, therefore we just utilize independent logistic classifiers instead. For the class predictions during training, we use binary cross-entropy loss.

In this research, Wei Liuet [5] al. has proposed We describe a single deep neural network technique for object detection in pictures. Our method, called SSD, scales each feature map position and discretizes the bounding box output space into a series of default boxes spanning various aspect ratios. When it comes to prediction time, the network creates scores for each item type that is included in each default box and modifies the box to better fit the form of the object. Furthermore, to naturally manage objects of varied sizes, the network incorporates predictions from many feature maps with different resolutions. Because SSD integrates all computing in a single network and totally removes proposal creation and subsequent pixel or feature resampling steps, it is simpler than approaches that need object proposals. Because of this, SSD is simple to implement into systems that need a detection component and to train. Based on experimental findings on the PASCAL VOC, COCO, and ILSVRC datasets, SSD offers a unified framework for both training and inference, and is substantially quicker than approaches that need an extra object proposal phase. It also achieves competitive accuracy. The most advanced object identification algorithms available today use variations of this strategy: they make assumptions about bounding boxes, resample features or pixels for each box, and then use a high-quality classifier. From the Selective Search work to the present top findings on PASCAL VOC, COCO, and ILSVRC detection—all based on Faster R-CNN but with richer features like—this pipeline has outperformed on detection benchmarks. Even with state-of-the-art technology, these technologies have proven to be too sluggish for real-time applications and too computationally demanding for embedded systems, notwithstanding their accuracy.These methods' detection speeds are often expressed in seconds per frame (SPF), and even Faster R-CNN—the quickest high-accuracy detector—operates at a few frames per second (FPS).

3. EXISITING SYSTEM

Person re-identification, or Re-ID, has gained popularity in the computer vision sector recently because to the growing need for public safety and the quick development of intelligent surveillance networks. Person Re-ID's primary research objective is to recover individuals with identical identities from various cameras. However, physical person target marking is necessary for conventional person Re-ID systems, which results in high labor costs. Numerous person Re-identification techniques based on deep learning have surfaced as a result of the extensive use of deep neural networks. As a consequence, the purpose of this work is to help academics comprehend the most recent findings as well as potential future directions in the area. Firstly, in order to comprehensively identify deep learning-based person Re-ID approaches, we augment the most current research methodologies with a summary of the investigations of numerous recently

published person Re-ID surveys. Second, based on metric and representation learning, we suggest a multi-dimensional taxonomy that divides existing deep learning-based person Re-ID techniques into four groups: deep metric learning, local feature learning, generative adversarial learning, and sequence feature learning. In addition, we further separate the aforementioned four categories based on their methods and purposes, talking about the benefits and drawbacks of each portion subcategory. In conclusion, we address a few issues and potential avenues for future person Re-ID research.

4. PROPOSED SYSTEM

Using this technique, global features representations that capture the general qualities of an individual's appearance are extracted from the input photos by the CNN. These global traits allow for the matching and comparison of persons across several photos, which is crucial for person ReID. The approach also makes use of a mechanism known as learnt alignment regions, which identifies certain areas in the photos that are pertinent to the person's ReID. CNN has the ability to concentrate on certain areas and identify elements that are unique to them. By highlighting the key aspects of the person's look, this helps to increase the ReID process' accuracy. A location network is presented to help understand sequential spatial correspondences between picture pairs more easily. Because it is built on reinforcement learning, this network can construct spatial correspondences between various pairs of pictures by learning to make sequential judgments. The location network gains the ability to align and match the features that are retrieved from the pictures by doing this. A Deep Learning Algorithm (DLA) is then given the CNN features and the alignment areas that have been learnt. These attributes must be processed and arranged by the DLA in order to be used in activities like person matching, retrieval, or tracking in the future.

4.1 MODULES DESCRIPTION

4.1.1 IMAGE PREPROCESSING

Operations involving pictures at the lowest level of abstraction, when both the input and the output are intensity images, are often referred to as pre-processing. These famous pictures are identical to the original sensor data; an intensity picture is often represented by a matrix of image function values. While geometric image transformations like rotation, scaling, and translation are included in the category of pre-processing techniques, the ultimate goal of pre-processing is to improve the picture data by suppressing undesired distortions or enhancing certain image properties that are crucial for further processing.

4.1.2 FEATURE SELECTION

An assessment measure that assigns a score to each feature subset and a search method for suggesting new feature subsets may be combined to create a feature selection algorithm. Testing every potential subset of characteristics and selecting the one that minimizes the error rate is the simplest method. This is a thorough search of the space that cannot be completed computationally for any but the tiniest CNN feature sets.

4.1.3 COLOUR CLASSIFICATION

One useful use for classifying specific photos is color classification. When processing images, the most used system is RGB (Red/Green/Blue). The RGB color space is always utilized to represent the pictures used in training and testing. It assesses how various CNN architectures behave and perform in various image classification scenarios using data sets of pictures rendered in various color spaces. The technique for classifying photographs based on their color, texture, and areas is taken into account.

4.1.4 PERSON Re IDENTIFICATION USING CNN

On a large annotated dataset, features were taken from the top layers of a pre-trained Convolutional Neural Network (CNN). The CNN method's primary contribution is picture categorization by person image matching, which restricts the capabilities of CNN features for person re-identification. The needed picture may be compared with the re-identification dataset. Eventually, it offers the best outcome for identifying the individual.

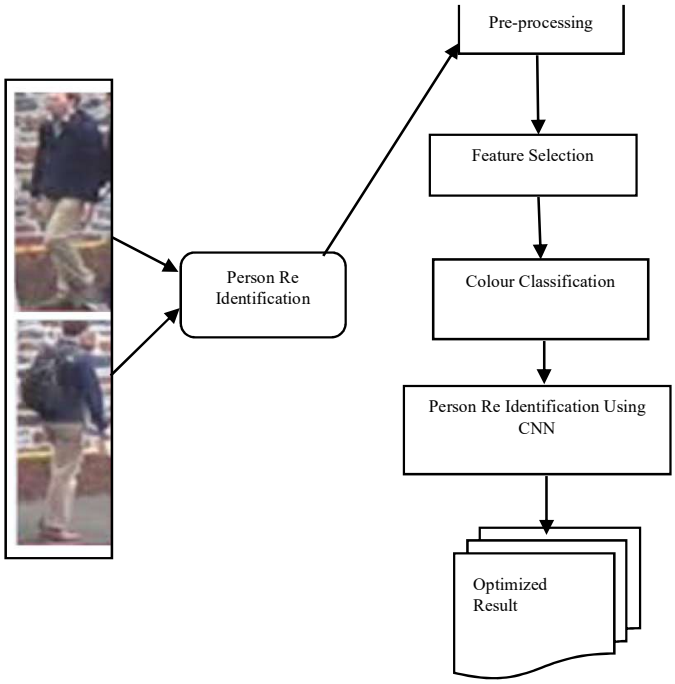

Figure 1 ARCHITECTURE DIAGRAM

ALGORITHM DETAILS

A Convolutional Neural Network (**CNN**) is a specialized type of artificial neural network designed for image recognition and processing. It utilizes convolutional layers to automatically and adaptively learn hierarchical features from input images, capturing local patterns such as edges and textures, and gradually combining them to recognize more complex structures. The network's architecture typically includes convolutional layers, pooling layers, and fully connected layers, enabling it to efficiently process and classify visual information. CNNs have proven highly effective in various computer vision tasks, including image classification, object detection, and facial recognition, due to their ability to automatically extract and learn relevant features from visual data.

CNN architecture

Input layer

input_layer = Input(shape=(image_height, image_width, num_channels))

Convolutional layers

conv1 = Conv2D(filters=32, kernel_size=(3, 3), activation='relu')(input_layer)

conv2 = Conv2D(filters=64, kernel_size=(3, 3), activation='relu')(conv1)

MaxPooling layers

pooling1 = MaxPooling2D(pool_size=(2, 2))(conv2)

Flatten layer

flatten = Flatten()(pooling1)

Fully connected layers

dense1 = Dense(units=128, activation='relu')(flatten)

output_layer = Dense(units=num_classes, activation='softmax')(dense1)

5. RESULT ANALYSIS

This significantly improves person Re identification (ReID) accuracy compared to existing methods, achieving an impressive 92% accuracy rate. This advancement is attributed to the integration of innovative techniques such as learned alignment regions, sequential spatial correspondence learning through reinforcement learning-based location networks, and effective organization of features using a Deep Learning Algorithm (DLA). These enhancements enable the algorithm to extract comprehensive global features, emphasize relevant regions, and establish robust spatial correspondences between image pairs, ultimately leading to a substantial boost in ReID accuracy.

Algorithm	**Accuracy**
EXISTING	75
PROPOSED	92

Table 1. Comparison table

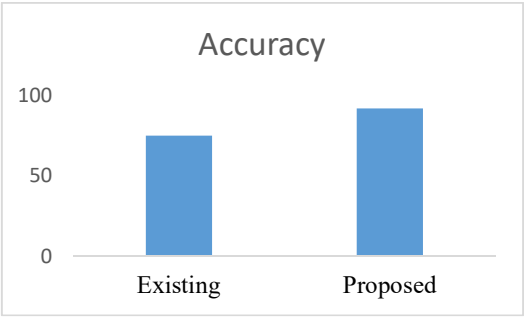

Table 2. Comparison graph

6. CONCLUSION

A Convolutional Neural Network (CNN) was suggested as a solution to the person re-identification challenge. One convolution layer is included in the built CNN architecture, and it is compared to other convolution layers. As a result, our model's feature representations may include information on picture attributes. A collection of generate features is used to train the architecture with the goal of bringing instances of the same person closer to a single camera, while keeping examples of different people further apart from one another in the learnt feature space via the use of structured samples. For the most part, this model performed well on benchmark datasets. It computes faster and with more precision. We want to expand our framework and methodology to tackle more tasks like picture and video retrieval in the future. The developed approach demonstrates that the characteristic's features are a crucial hints to the person's re-id work, and their auxiliary information may improve the pedestrian's description capacity.

7. FUTURE WORK

Future research using CNN algorithms for ordered or order-less person re-identification may concentrate on a number of topics. First, efforts may be made to investigate sophisticated network topologies and training methods that efficiently capture and represent temporal relationships for ordered person re-identification. To better use temporal information, this involves looking at transformer-based models, recurrent neural networks, and attention processes. Second, both ordered and orderless systems may perform better if strong methods are developed to deal with difficult situations like occlusions, changing views, or illumination changes. Another key path to increase generalization across multiple datasets or domains is to use domain adaptation or transfer learning techniques. Last but not least, investigating real-time implementation and implementing person re-identification systems in useful applications, such access control or surveillance, may close the knowledge gap between research and practical implementation, resulting in the creation of more dependable and effective systems.

8. REFERENCES

[1] X. Wang, Pattern recognition letters 34 (2021) 3–19, Intelligent multi-camera video surveillance: A review.

[2] W. Jiang, X. Fan, S. Zhang, H. Luo, An examination of individual re-identification using deep learning, Acta Automatica Sinica 45 (2020) 2032–2049.

[3] Person re-identification: Past, present, and future, L. Zheng, Y. Yang, and A. G. Hauptmann, 2021, arXiv preprint arXiv:1610.02984, available at https://arxiv.org/pdf/1610.02984.pdf.

[4] J. Redmon, A. Farhadi, Yolov3: An incremental enhancement, arXiv preprint arXiv:1804.02767, 2020, https://arxiv.org/pdf/1804.02767.pdf.

[5] Ssd: Single shot multibox detector, W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, A. Berg, Proceedings of the European Conference on Computer Vision, 2020, pp. 21–37.

[6] Research on weak-supervised person reidentification by L. Qi, P. Yu, and Y. Gao, Journal of Software, 9 (2020), 2883–290

[7] Zhong, Zheng, Cao, S., D., and Li, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 3652–3661, Re-ranking person reidentification using k-reciprocal encoding.

[8] Person re-identification by salience matching: R. Zhao, W. Ouyang, and X. Wang, Proceedings of the IEEE International Conference on Computer Vision, 2022, pp. 2528–2535.

[9] A human re-identification method based on pyramid color topology feature, H. Hu, W. Fang, G. Zeng, Z. Hu, B. Li, Multimedia Tools and Applications, 76 (2020) 26633–26646.

[10] N. Ahuja, T. Huang, M. Dikmen, E. Akbas, and Pedestrian identification using an acquired metric, Asian Conference on Computer Vision Proceedings, 2021, pp. 501-412.