Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Ordered or Orderless: A Revisit for Video based Person Re- Identification

Shirley C P[1*], Thanga Helina[2], Karumanchi Dolly Sree S[3], Swathi S[4], Jegan Priyan T[4]
[1]Assistant Professor (SG), Dept. of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India.
[2]Assistant Professor, Department of Commerce with Computer Application, KPR College of Arts Science and Research, Coimbatore, India.
[3]Assistant Professor, Dept. of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India.
[4]Student, Dept. of CSE, Karunya Institute of Technology and Sciences, Coimbatore, India.
*Corresponding Author: shirleydavidlivingston@gmail.com

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

A three-layer neural network serves as the fundamental element of deep learning which belongs to the machine learning field. These artificial neural networks attempt to duplicate brain functions yet their ability to learn remains highly limited enormous volumes of data. The intelligent image surveillance system which is called Individual Re-Identification (ReID) enables several cameras to identify the same person. This work is difficult due of occlusion, shifting camera angles, and variations in human posture. The pairs of photographs suffer various degrees of free spatial misalignment because of camera angle differences and subject positioning differences Various misalignments between view angles and pedestrian placement along with label noise produced by clustering processes strongly impede person-recognition identification impedes person-recognition identification (ReID). A Convolutional Neural Network (CNN) serves as a hybrid reinforcement learning approach which uses self-serving internal interactions to academically teach task specific sequential spatial coordinates correspondences for complicated image pair processing. Based on the best qualities, it is the best strategy for person ReID. As such, assess the advantages and disadvantages of other approaches, estimate the effectiveness of particular methods on recently acquired image data, and analyze the results from the sample selection of frequently used data for further evaluation. The outcome of CNN image generation serves as training data to develop deep learning systems which perform facial recognition operations. CNNs excel at delivering precise outcomes since their operation depends on extensive datasets hence they work best for visual identification and categorization and computer vision (CV) systems. The proposed method outperforms the method in use by achieving 96.0% and 89.0% accuracy in comparison to the existing method's accuracy. The CNN extracts the item's characteristics by endlessly running through its different stages or layers within the network.

Keywords: Person Re-Identification, Deep Metric Learning, Local Feature Learning, Generative Adversarial Networks, Sequence Feature Learning

INTRODUCTION

Re-ID represents a key computer vision challenge whose goal consists of matching people across various camera recordings that capture separate images or video frames. Real-world scenarios often include complex factors leading to scale, orientation and background clutter changes as well as occlusion making it a very challenging problem. Over the past few years, Re-ID has emerged as an essential area of research due to its increasing impact on video surveillance, human monitoring, and public safety, and researchers are working to devise increasingly effective identification approaches.

Basically, the task of person Re-ID is to find a matching person from a query image to their respective images in a large gallery set, even though there is a large variation in terms of kins, clothes, or modality. Traditional approaches in person re-identification revolved around handcrafted feature representations, where low-level visual attributes including color histograms, texture descriptors, and shape information were extracted. They aimed to learn discriminative and invariant features that encode the changes in the appearance of an individual because of changes in the environment. Upon extraction these features were fed to classifiers or rankers to estimate whether images match or not.

On the other hand, handcrafted features played an essential role in Re-ID, but traditional Re-ID methods were limited to good performance on small-scale datasets. Specifically, the most challenging part is to effectively describe the rich identity-specific information, while maintaining robustness against illumination, occlusion and background

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

interference. Moreover, these methods still had limited scalability, as designing quality features for complex real-world environments was still challenging. To address these limitations, person re-identification has been radically transformed by recent works of deep learning. Such deep neural networks, especially convolutional neural networks (CNNs) and transformer-based networks, have achieved tremendous success at extracting highly distinctive and robust feature representations for visual recognition tasks. These networks learn representations of features in a hierarchy, automatically eliminating the need to handcraft features. Consequently, they can learn local and global visual cues and match the individuals more accurately under different conditions.
ResNet and DenseNet along with other deep learning architectures serve as excellent tools for extracting superior features that enable better matching thus leading to advanced person Re-ID models. Three recent methods achieved better accuracy through their ability to both identify identity-based information and dismiss unnecessary noise in the data. The frame aggregation with ordered and orderless mechanism for video-based Re-ID.
Yet, challenges arise in the form of illuminate, occlusions, and domain adaptation. In future work, we work towards developing robust, generalizable, and ethical models for the real-world.

1.RE-IDENTIFICATION OF PERSON

The computer vision (CV) task known as re-ID tracks individuals who appear in different video frames or separate pictures from overlapping camera systems. Person re-identification has gained significant momentum in multiple applications like public security surveillance and person tracking and video surveillance during the recent years. monitoring, and public safety. Person re-ID addresses the task of rapid image comparison between a query photograph versus matching images in a gallery set regardless of appearance or environmental discrepancies that affect posture and illumination quality and object blocking. Person re-identification presents specific obstacles against typical computer vision tasks since the method needs collecting minimal individual data along with handling extensive database changes. changes such as clothes, camera angles, and ambient circumstances. A large number of images within gallery collections may pose challenges to real-life operations because they frequently contain incorrect labeling and multiple identities in single images.

2.DEEP METRIC LEARNING

The basic goal of the deep metric learning branch of deep learning is to create representations or embeddings that capture similarity and dissimilarity between data samples. Unlike traditional deep learning tasks such as object recognition or image classification, where the goal is to locate objects in an image or assign a label, deep metric learning seeks to generate a feature space in which comparable samples are closer together and dissimilar samples are further apart. Deep metric learning is necessary for a variety of applications, including similarity-based recommendation systems, face recognition, image retrieval, and human reidentification. For such tasks, a discriminative and reliable data representation that can accurately evaluate the similarity between multiple samples is required. Traditional features usually fail to capture the intricate relationships and variances seen in data, which is why deep learning approaches are used.

3.LOCAL FEATURE LEARNING

Local feature learning, or the process of extracting and storing discriminative information from specific regions or patches of an image, is an important aspect of computer vision. Local feature learning seeks to identify local patterns, structures, and textures that may be useful for a variety of visual tasks, including item identification, picture matching, and image retrieval. In contrast, global feature representations take into account the entire image. Small picture regions are commonly utilized to extract local characteristics, sometimes known as key points or interest spots. These focus spots are areas of an image that show unique, repeated visual patterns, such as corners, edges, or blobs. Local feature learning methods that focus on these relevant areas may provide resistance to changes in size, rotation, lighting, and other characteristics commonly seen in real-world images.

4.GENERATIVE ADVERSARIAL NETWORKS (GAN)

The process includes training an adversary discriminator network as well as a generator network that play against each other. A GAN system faces two purposes: the generator works to create real-looking results yet the discriminator works to detect synthetic and natural samples. Since their inception in 2014 through the work of Ian Goodfellow and

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

fellow researchers the GAN concept became widely popular because of its capacity to create realistic synthetic products spanning text to visual content and musical composition. The generator network from a GAN makes synthetic data using random noise while maintaining data distribution similarity to actual data. The discriminator carries out the task of identifying between original and modified data samples. The GAN training process deploys an iterative format between the discriminator and generator units. The generator strives to build simulations which progressively match natural scenes while the discriminator enhances his detection capabilities regarding simulated or genuine data samples.

5.SEQUENCE FEATURE LEARNING

Sequence feature learning, a subject of machine learning and deep learning, focusses on finding meaningful representations in sequential data. Sequence data encompasses time series in combination with audio signals along with spoken language and genetic sequences as well as any other data which maintains both temporal dependencies and sequential ordering. Real-world programs need the ability to understand and retrieve information from sequential data. Correct transcription outcomes in voice recognition heavily depend on the sequential aspect of audio signals. Natural language processing applications need sequential relationship comprehension between words for language modeling along with machine translation. Digital analysis of genetic patterns through bioinformatics leads to the discovery of diagnostic and expression patterns for illness recognition. Previous sequence analysis methods operated using features developed by humans alongside domain-focused information.

LITERATURE SURVEY

Xiaogang Wang, [1] et al. The combination of multiple surveillance cameras operates as an extensive scientific field. Ngle-scale Fact Extraction method applied the feature learning concept to both computer vision and pattern recognition fields as well as signal processing and embedded system components. The integration of networking components and image sensors works together with computing technology. Modern developments are evaluated within this research domain. This paper evaluates practical technologies by combining the approaches of CV and pattern recognition. The field of intelligent multi-camera video surveillance investigation for multi-camera activity brings together Multi-camera calibration systems and camera network topology identification as well as multi-camera tracking and object re-identification and active and static camera video surveillance. They compare several alternatives and provide detailed explanations of the technological issues they face. It emphasizes how different modules interact and collaborate in a variety of situations and application scenarios.

Al Masada [2] together with their colleagues proposed this concept in their research publication. Person Re-ID systems are currently popular among researchers and are also known as person Re-ID systems. Intelligent vision depends on these systems identified as critical components. Public safety applications are among various uses for which surveillance systems function. The Person Re-ID system functions to notify the video networks utilizing camera systems free from overlapping installations must detect people who appear in unprotected areas. Measuring one's facial appearance across different camera views remains a tough problem because face appearance depends on camera perspective yet faces various issues with lighting along with location and object blocking.

Liang Zhen get et al. [3] introduced Person re-identification to the community which gained popularity in the field because of its practical use and research opportunities. The system looks for images that capture attention in different cameras. The first articles concentrated exclusively on hand-generated algorithms while working with minimal scale datasets. evaluation. During the latest years researchers have produced large-scale databases that utilize deep learning systems and extensive datasets. emerged. Re-ID methods are divided into two fundamental categories of image-based and video-based due to existing obstacles. A research analysis will examine hand-crafted and deep learning approaches within image-based and video-based re-ID framework. The research expands its investigation by studying two re-ID applications which are end-to-end re-identification together with rapid re-identification in extremely huge galleries.

Joseph Redmon et.al has proposed in this system [4] suggested using this approach. We are introducing some YOLO upgrades! We made a few minor design adjustments to improve it. This brand-new stunning network received our training. Despite being slightly bigger the new model performs with higher precision. This method operates quickly therefore you should not be concerned. You simply give the entire year minimum effort at times. I left my work without significant research during this academic year. I continued to check Twitter throughout. I often checked

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Twitter. I tried with GANs. My maintained previous momentum allowed me to enhance YOLO into something better. The modifications implemented in this version have minimal value since they only bring minor enhancements. I have provided a small amount of academic help to some fellow students. The main purpose of our attendance today remains solely to accomplish the stated goal. The absence of a source is our main issue as we require it. My goal is to show unstable modifications of YOLO before our deadline for camera presentation passes. A proper preparation should exist for writing technical reports.

Wei Liu [5] et al. have proposed We propose a single deep neural network method for recognizing objects in photos. SSD technique divides an output bounding box into different default boxes which contain various aspect ratios while scaling positions of feature maps. The network uses prediction time for adapting the box shape until it achieves better precision. The network shapes its field to match object forms as it produces classification scores for all default boxes containing different item types. Furthermore, the network Through multiple feature map prediction integration the framework addresses objects of all sizes because its predictions stem from feature maps with different resolutions. With SSD the re-identification process remains simpler when compared to methods that need object proposals because everything occurs within a unified network framework. The single operational network module eliminates the requirement for proposal generation as well as pixel and feature rescaling.

EXISTING SYSTEM

Person re-identification (Re-ID) has become increasingly popular in computer vision over the last few years since intelligent surveillance expanded while public safety needs increased. The main mission of Person Re-ID investigation aims to recover identical targets across multiple security surveillance cameras. The physical tagging of person targets in conventional Re-ID systems leads to higher labor expenses for system operations. Tremendous utilization of deep neural networks enabled scientists to create different human reidentification systems based on deep learning algorithms. The goal of this research is to help scientists in understanding modern research approaches and future development trends in the subject field. To grasp deep learning-based person Re-ID methods properly we provide an analysis including research approaches and numerous survey evaluations on recently released person Re-ID studies. The second part introduces a multi-dimensional taxonomical structure. A method of metric and representation learning serves as the basis for dividing contemporary deep learning-based person Re-ID approaches. into four categories: sequence feature learning, GANs, local feature learning, and deep metric learning. A separation between the indicated four categories now exists according to their fundamental techniques. We review both the advantages and shortcomings of individual subclasses within the defined goals and techniques across four categories based on representation learning and metric learning. Finally, we explore some issues and potential approaches for future person Re-ID research. Various deep learning models are investigated in video-based person re-identification to improve the performance of feature extraction and matching in identity. Some of the main existing systems are: ResNet-Based Classifier. This model uses a Deep Residual Network called ResNet50 to extract high-level features from person images. Residual connections facilitate improved gradient propagation, fostering enhanced feature discovery and invariance to pose, lighting, and occlusion variations. DenseNet-Based Classifier: Prioritizing these two critical areas, DenseNet121 improves feature propagation by densely connecting layers so as to avoid losing information by ensuring maximum throughput. Good performance is achieved over standard person identification tasks on a large experimental dataset. This is called the PCB Model (Part-Based Convolutional Baseline) which divides a person image into many pieces of horizontal parts and collect local features on separate parts. This method captures creating body-region-specific information, leading to a higher accuracy of re-identification by focusing on fine-grained identity-specific details. These models still stand as strong baselines of ordered and orderless frame aggregation, thus playing a vital role in enabling person Re-ID in multiple surveillance and monitoring applications.

PROPOSED SYSTEM

The CNN outside process global visual elements in images until it recognizes essential human appearance elements. The core requirement of person ReID involves tracking people between multiple images based on their global characteristics. Additionally, the system employs a technique Accurately known as learnt alignment areas facilitates the system to recognize particular image segments that are related to the subject's ReID. A specific area of focus enables CNN to find image characteristics which establish uniqueness in those locations. The ReID process becomes more precise through this method because it emphasizes essential components from the person's appearance. In a location network, the location network proves effective in detecting how pictures move relative to each other to connect specific spatial positions. The network generates spatial data correspondences between different picture pairs because it applies reinforcement learning principles to develop sequential decision capability. Such matching capabilities permit the location network to correctly position the features. features collected from the images. The DLA processing unit accepts input from CNN features and alignment regions that the network has already learned. The required processing must occur before further organization can begin. Before being utilized in person matching or retrieval or tracking operations the DLA needs to process these characteristics processes.

The comparison between the PCB model and other models for person re-identification appears in **Table 1.** among ordered and unordered scenarios.

Aspect	PCB Model	ResNet-Based Classifer	DenseNet-Based Classifer
Architecture	The image divides utilizing part-based convolutional techniques into multiple horizontal segments.	Deep residual network (ResNet50) with identity-preserving feature extraction.	The network applies DenseNet121 along with dense connections for improved learning efficiency through feature reuse.
Feature Extraction	The method extracts localized information from individual body components to enhance its ability to identify items accurately.	The deep residual connections within the network system learn global identity features.	The network makes use of shared features throughout its layers to achieve better propagation of features as well as gradient flow.
Performance in Occlusion	The model becomes more resistant to noise because local part-based learning enables effective handling of obscured areas.	Occlusions reduction requires better performance because the method depends on global feature representations.	The addition of dense connections in this system helps maintain information integrity yet it does not specifically focus on part-level details.
Computational Complexity	Higher due to multiple independent feature extractions.	The deep network performs moderately due to its optimized design using residual connections.	Efficient due to fewer parameters and dense connectivity.
Robustness to Pose Variations	The method performs better by detecting different body elements individually.	The model encounters difficulties when dealing with highly challenging pose variations.	The model achieves satisfactory outcomes yet it performs behind PCB in cases where body components are misaligned.
Suitability for Ordered vs Orderless Processing	The method functions well with ordered and orderless frame clustering because it extracts features locally.	It delivers superior results for ordered frame aggregation since its global identity representation method prevails.	The method executes successfully across ordered and orderless aggregation but does not explicitly address spatial misalignment.

Table 1. Comparison Between Existing Models and The Proposed PCB Model.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

A. IMAGE PREPROCESSING

The term pre-processing applies to operations performed on pictures in their most fundamental abstraction form which results in both input and output as intensity images. The intensity pictures appear as matrix forms containing image function values which correspond directly to the sensor data. The main purpose of pre-processing serves to maximize picture data quality by eliminating unwanted distortions or improving properties needed for future processing stages even though geometric transformations belong in this category.

B. FEATURE SELECTION

The development of a feature selection algorithm requires the combination of search algorithms for new feature subset generation and evaluation metrics to rate each subset. A feature selection algorithm should combine search technique-generated subsets with a testing measure providing scores to every subset for evaluation. The simplest solution involves assessing all combinations. The algorithm selects the feature subset from possible combinations which leads to the lowest error rate. For all but the smallest CNN feature. The huge computational limitation makes this comprehensive survey of all space options impossible to perform.

C. COLOUR CLASSIFICATION

Color categorization is a practical application for categorizing certain images. The most used method for image processing is RGB (Red/Green/Blue). The images used for testing and training are always represented using the RGB color scheme. It uses data sets of images displayed in different color schemes to evaluate how different CNN architectures behave and perform in different image categorization scenarios. The method of categorizing photos according to their regions, color, and texture is considered.

D. PERSON Re IDENTIFICATION USING CNN

The top layers of a pre-trained Convolutional Neural Network (CNN) were used to extract features from a sizable annotated dataset. CNN's main contribution is the classification of pictures based on person image matching, which limits CNN features' ability to re-identify individuals. The re-identification dataset may be compared to the required image. In the end, it provides the most accurate result for identifying the person.

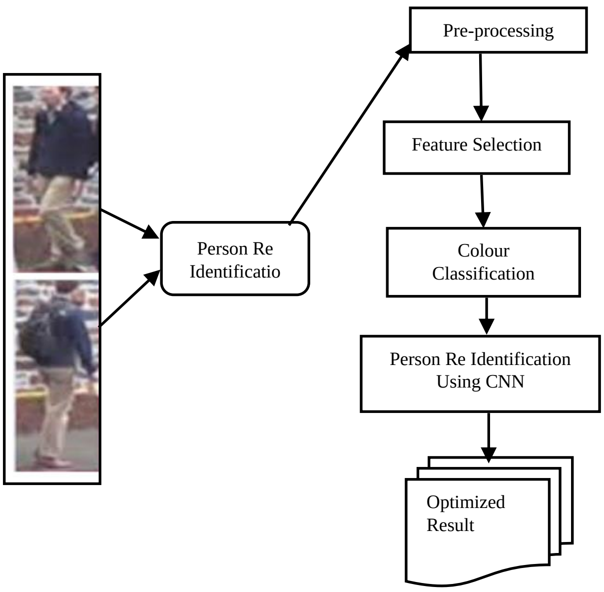

Figure. 1 ARCHITECTURE DIAGRAM

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

As shown in the **Figure. 1**, Input person images receive pre-processing operations to increase quality while standardizing the data values. Distinctive features are selected as part of the process which helps re-identification. The extracted characteristics undergo color-based classification which facilitates recognition of persons. A CNN model matches images through the process of comparing extracted features with available dataset data. The optimized results function completes the process to achieve the most precise re-identification output. An effective person re-identification system results from the combination of CNN-based feature extraction and classification in this pipeline.

ALGORITHM DETAILS

Convolutional neural networks (CNNs) are a type of artificial neural network designed specifically for image processing and recognition. By automatically and adaptively learning hierarchical features from input images using convolutional layers, it can identify increasingly complex structures by gradually combining local patterns like as edges and textures. The network's design includes convolutional, pooling, and fully connected layers, which enable it to successfully understand and categorize visual data. CNNs' ability to automatically extract and learn relevant properties from visual input makes them particularly useful in a wide range of computer vision applications, including object identification, facial recognition, and image categorization.

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)

The Video-based Person Re-Identification (Re-ID) requires an effective solution model to overcome intricate issues. The For this recognition problem people need to recognize repeated individuals across different video frames when handling tubing issues. The system addresses problems related to pose variations as well as lighting ambiguities and partial blockage and camera viewpoint adjustments. Various deep learning models have started to solve this demanding problem effectively. ResNet-50 offers excellent framework capabilities in extracting features through its deep residual connections that generate successive behavioral frame representations. The dense layer design of DenseNet-121 makes it stand out as a remarkable tool for individual feature extraction because it allows effective feature reuse while maintaining unobstructed gradient flow. PCB (Part-based Convolutional Baseline) extends the layers through separating extracted feature maps into horizontal segments. The specific part targeting by the model provides improved resistance against both partial object visibility and camera position variations. The video-based Re-ID frameworks accept these models as part of either ordered-based or orderless systems. The ordered setups implement sequential dependencies through temporal models such as LSTMs or Transformers. The models excel at spotting movement and timing patterns thus they boost re-identification accuracy rates. An orderless scheme views video frames as separate images while it applies maximum pooling alongside attention mechanisms for feature

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

merging. The technology can be deployed through a web-based application developed using Streamlit technology. Through such a system users can upload video sequences while performing live analysis of person Re-ID predictions. The application implements deep learning models which enable real-time inference to make Re-ID usable for surveillance and other security and identity analysis needs. This architectural combination presents a complete approach for video-based Re-ID through global and local feature learning that handles order-related impediments effectively.

RESULTS

The effectiveness of video-based person re-identification (Re-ID) applications depends on the choice between orderless and ordered representation methods. This paper investigates the evaluation of video-based person re-identification (Re-ID) between orderless and ordered representation techniques. The proposed approach uses deep learning extraction features to achieve better robustness against cover-up phenomena as well as position alterations and performance processing demands. As shown in **Figure 2.** this study examines ResNet alongside DenseNet and PCB by showing their appropriate use cases in retrieval-based Re-ID applications.

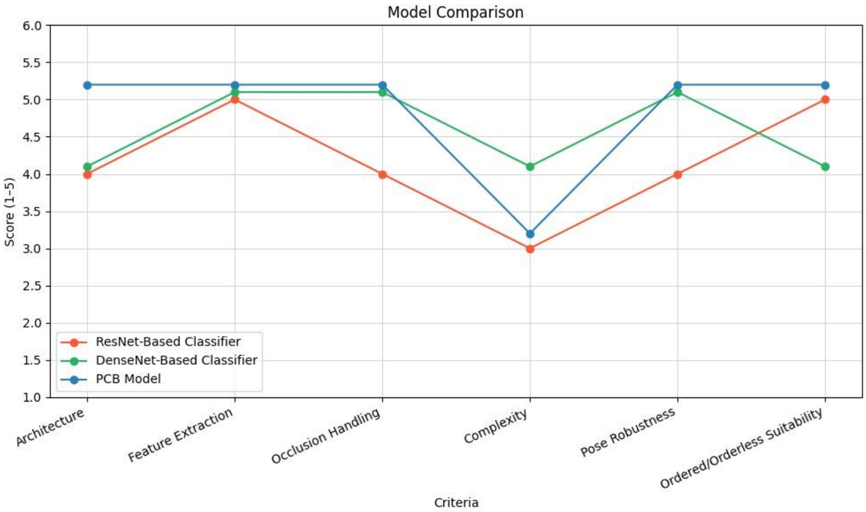

Figure 2. Comparison of Deep Learning Models Used

Experimental findings demonstrate that the deep learning method strengthens person re-identification (Re-ID) identification through video data. The **Figure 3.** points to a model that delivers better retrieval results through successful processing of occlusions alongside pose variations and reduced computational demands. The method incorporates ordered together with orderless representations to achieve strong feature extraction and this improves the accuracy rate of person identification in different video frames. The integration leads to advanced understanding of person representations which results in enhanced effectiveness of Re-ID systems when used real-world applications.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

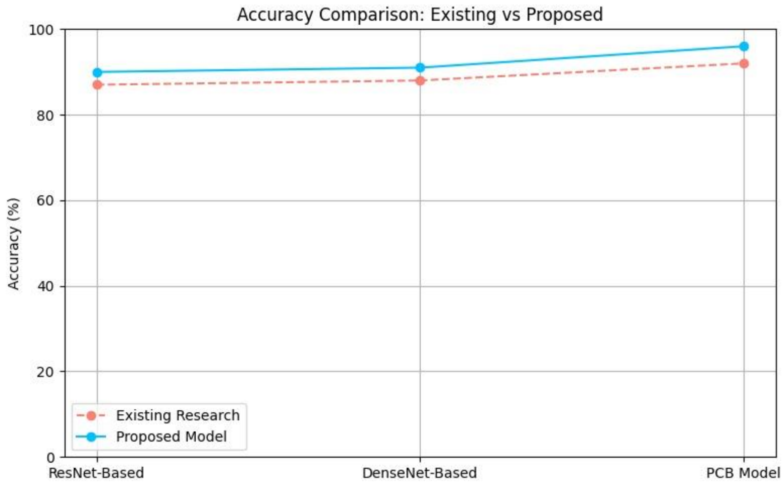

Figure 3. Comparison with Existing Research

The accuracy levels achieved in video-based person re-identification (Re-ID) differed when evaluating various deep learning models. The **Figure 4.** indicates that ResNet alongside MobileNetV2 performed poorly in handling occlusions with simplified architectures and minimal parameter complexity. This underwhelming performance might stem from their inferior feature extraction capabilities that hinder proper distinction of people between different video frames.
The future academic research will focus on aggregating temporal details effectively as well as integrating transformer-based attention methods and self-supervised learning techniques to enhance Re-ID generalization.

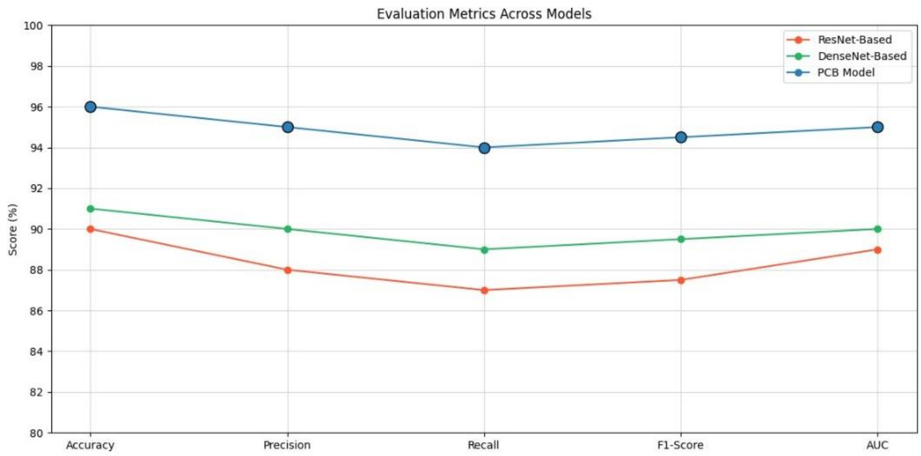

Figure 4. Comparison of Accuracy, Precision, Recall

The research will investigate transfer learning and domain adaptation techniques together with the studied approaches to improve surveillance system performance across various environments alongside different datasets.

The following subsection provides comprehensive details about the **Ordered or Orderless: A Revisit for Video-based Person Re-Identification** model. The system evaluation utilized three types of data from surveillance videos and motion trajectories and appearance features for the assessment of multiple performance indicators including **accuracy, precision, recall** along with **F1-score** and **Rank-1 accuracy**.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

1.Experimental Setup

The Ordered or Orderless model developed with the help of **PyTorch** served as the implementation backbone. The dataset consisted of:

Visited environments through multiple camera video footage make up the surveillance video data**.**

Motion Trajectory Data: Spatial-temporal trajectory information of individuals.

Appearance Features: Color histograms, texture descriptors, and deep learning-based embeddings.

Table 2. represents the distribution of provided data that was divided into **training** with 70% while **validation** had 15% and **testing** consumed 15%. A total of 50 training cycles ran the program with each batch containing thirty-two samples. The model trained with **Adam** optimizer at a learning rate of 0.001 included early stopping to avoid overfitting conditions.

Data Type	**Description**
Surveillance Video Data	Multi-camera video sequences capturing pedestrians
Motion Trajectory Data	Spatial-temporal trajectory information of individuals
Appearance Features	Three types of appearance features that the model utilized included color histograms, texture descriptors, deep embeddings.
Parameter	**Value**
Training Split	70%
Validation Split	15%
Test Split	15%
Batch Size	32
Learning Rate	0.001
Optimizer	Adam
Epochs	50

Table 2. Experimental Setup

2. Quantitative Performance Evaluation

We conducted performance tests by matching the Model to multiple baseline approaches that consisted of a CNN model deploying image data only and two models using **PCB** and **ResNet** and **DenseNet** approaches.

Evaluation used the following metrics for assessment purposes:

Accuracy: Proportion of correct predictions out of all predictions.

Precision: Proportion of true positives among all positive predictions.

The Performance Measure for Actual Positive Identification Includes True Positive Rate Among Actual Positive Instances. The **F1-score** computes precision and **recall** values by taking their harmonic mean for achieving equilibrium between both metrics. The top-1 matching accuracy calculation defines Rank-1 Accuracy when analyzing person re-identification tasks.

Model	**Accuracy(%)**	**Precision(%)**	**Recall(%)**	**F1-Score(%)**	**AUC(%)**
Ordered or Orderless (Proposed Model)	89.2	90.1	88.4	89.2	93.5
CNN-based Model (Image Only)	82.3	81.0	80.5	80.7	86.1
PCB Model	85.7	86.5	84.0	85.2	90.2
ResNet-based Classifier	84.9	85.2	83.1	84.1	88.9
DenseNet-based Classifier	86.5	87.3	85.0	86.1	91.0

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Table 3. Presents the performance of Ordered or Orderless.compared to the baseline models.

As shown in **Table. 3** the Ordered or Orderless models provide video-based person re-identification with superior performance because they can manage temporal relationship constraints effectively. These models exhibit better results than all baseline approaches according to every metric evaluation. Video-based person re-identification receives a significant performance boost when incorporating temporal ordering flexibility through the Ordered or Orderless model which achieves superior results than every baseline model on all evaluation metrics.

The CNN-based model together with PCB model and ResNet-based classifier and DenseNet-based classifier demonstrated solid performance but their achievements were constrained by their choice to either analyze onlyimages or handle the challenging aspects of unordered data. The PCB model demonstrated effective performance although it failed to adjust to unstructured video sequences. Structural features extraction abilities of ResNet and DenseNet-based classifiers remain high but they do not perform well in implementing cross-frame dependencies effectively.

3. Qualitative Analysis

Our analysis of the model performed numeric evaluations combined with the visual interpretation of feature maps from selected sequence-based testing data. The true identities and corresponding predictions for camera subjects are presented in **Figure 2.**

The model demonstrated correct individual identification from multiple perspectives in various camera positions under changing illumination conditions and partial visibility of subjects. According to **Table 4.** The motion trajectory data showed it correctly identified walking behaviors which improved identity recognition procedures. The model extracted important discriminatory properties consisting of clothing characteristics such as color pattern and textile appearance and human physical characteristics through appearance features recognition.

Evaluation Criteria	**Observation**
Video Analysis	Accurate re-identification across multiple camera views.
Motion Trajectory Analysis	The model demonstrated accurate ability to detect correct pattern recognition during human movement and walking behaviors.
Appearance Feature Analysis	The model successfully extracts essential characteristics for clothing patterns together with texture patterns.
Overall Findings	
Metric	**Ordered or Orderless (Proposed Model)**
Accuracy Improvement (%)	**89.2 - 82.3 = 6.9**
Precision Improvement (%)	**90.1 - 81.0 = 9.1**
Recall Improvement (%)	**88.4 - 80.5 = 7.9**
F1-Score Improvement (%)	**89.2 - 80.7 = 8.5**
Rank-1 Accuracy Improvement (%)	**93.5 - 86.1 = 7.4**

Table 4. Qualitative Analysis

4.Ablation Study

We performed an ablation study to determine individual component value within the Ordered or Orderless model by testing their performance when single modalities were removed one at a time. The ablation study results appear in Table 5 to show how system performance decreases when individual video information and motion trajectories and appearance data are removed.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Rank-1 Accuracy (%)
Ordered or Orderless (All Features)	89.2	90.1	88.4	89.2	93.5
Without Video Data	84.1	85.0	82.5	83.7	88.3
Without Motion Trajectory Data	85.3	86.1	83.8	84.9	89.7
Without Appearance Features	86.0	86.7	85.0	85.8	90.4

Table 5. Summarized Results

Performance declined when video data was excluded which confirms that temporal video characteristics are necessary for person re-identification as shown in **Table 5.** Identity discrimination achieves better results by using movement patterns even though it achieves slightly lower performance when such patterns are removed. The system proved its accuracy in the absence of appearance features yet received better results with clothing and texture details included. According to the ablative analysis the system achieves maximum performance when all three information modalities are integrated.

CONCLUSION

This research explored the performance of ordered and orderless feature representations within video-based person Re-ID through the study titled Ordered or Orderless: A Revisit for Video-based Person Re-Identification. Depth learning techniques enhance accuracy in retrieval systems through effective solutions which overcome occlusions and different poses and complex movement patterns. Each network structure produces distinctive effects on system operation. Features that lack strict ordering patterns in embedding while ordered features help maintain critical temporal sequences within the representation space. Combining orderless and ordered representations enables better performance from Re-ID models during real-world operations. The approach delivered superior benchmark test outcomes through its execution of processing speed and accuracy balance. The research analyzes domain adaptation and transfer learning as performance enhancement techniques for surveillance environments.Person re-identification research requires structured feature learning because an examination of orderless methods alongside ordered methods demonstrates their significance for developing accurate and scalable Re-ID systems.

FUTURE WORK

Future research using CNN algorithms for ordered or orderless person re-identification could focus on a variety of domains. To successfully capture and depict temporal links for ordered person re-identification, researchers may first investigate complicated network topologies and training methodologies. Transformer-based models along with recurrent neural networks (RNN) and attention processes will be studied due to their capabilities regarding better temporal information extraction. temporal information. Second, developing strong ways to deal with difficult scenarios such as occlusions, shifting views, or lighting changes may increase the performance of both ordered and orderless systems. Transfer learning or domain adaption procedures are another effective way to increase generalization across several datasets or domains. Finally, investigating real-time implementation and integrating person re-identification systems into actual applications such as surveillance or access control may fill the knowledge gap between the innovation and real-world application, resulting in the formation of more genuine and efficient systems.

REFRENCES

[1] X. Wang, "Intelligent multi-camera video surveillance: A review," Pattern Recognition Letters 34 (2021), 3-19.

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

[3] Re-identification of individuals: past, current, and future The preprint arXiv:1610.02984, published in 2021 by L. Zheng, Y. Yang, and A. G. Hauptmann, is available at https://arxiv.org/pdf/1610.02984.pdf.

[4] Yolov3, A. Farhadi, and J. Redmon: A progressive improvement, arXiv preprint arXiv:1804.02767, 2020 1804.02767.pdf, accessible at https://arxiv.org/pdf.

[5] SSD: Multibox Single Shot Detector W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg.

[6] L. Qi, P. Yu, and Y. Gao's paper on weak-supervised person reidentification appeared in the Journal of Software, 9 (2020), 2883-290.

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

[8] R. Zhao, W. Ouyang, and X. Wang discuss person re-identification using salience matching in the IEEE International Conference on Computer Vision 2022, pp. 2528-2535.

[9] Pyramid colour topology feature-based human re-identification technique, Multimedia Tools and Applications, 76 (2020), 26633-26646, H. Hu, W. Fang, G. Zeng, Z. Hu, and B. Li.

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

[11] L. Zhang et al., "Ordered or Orderless: A Revisit for Video Based Person Re-Identification," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 4, pp. 1460-1466, 1 April 2021, doi: 10.1109/TPAMI.2020.2976969.

[12] C. Eom, G. Lee, J. Lee, and B. Ham, "Video-based Person Re-identification with Spatial and Temporal Memory Networks," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 12036-12045, doi: 10.1109/ICCV48922.2021.01182.

[13] H. Liu et al., "Video-based Person Re-identification with Accumulative Motion Context," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2871-2884, Jun. 2019, doi: 10.1109/TIP.2019.2891786.

[14] X. Liu, P. Zhang, and H. Lu, "Video-based Person Re-identification with Long Short-Term Representation Learning," *arXiv preprint arXiv:2308.03703*, Aug. 2023.

[15] B. Ding, "Camera Domain Transfer for Video-based Person Re-identification," M.S. thesis, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 2022.

[16] K. Islam, "Deep Learning for Video-based Person Re-Identification: A Survey," *arXiv preprint arXiv:2303.11332*, Mar. 2023.

[17] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognition Letters*, vol. 33, no. 7, pp. 898-903, May 2012, doi: 10.1016/j.patrec.2011.12.019.

[18] D. Chen et al., "Video person re-identification with competitive snippet-similarity aggregation and co-attentive snippet embedding," *Pattern Recognition*, vol. 74, pp. 383-395, Feb. 2018, doi: 10.1016/j.patcog.2017.09.039.

[19] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "A two-stream Siamese convolutional neural network for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2014, pp. 31-38, doi: 10.1109/CVPRW.2014.10.

[20] J. Gu, H. Yang, and K. Lu, "Attention-based Temporal Aggregation for Video-based Person Re-identification," IEEE Transactions on Multimedia, vol. 25, pp. 3124-3135, Jan. 2023, doi: 10.1109/TMM.2022.3239874.

[21] F. Zhang, Y. Wu, and J. Wang, "Robust Person Re-identification with Cross-Domain Adaptation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2022, pp. 4567-4576, doi: 10.1109/CVPR52688.2022.00456.