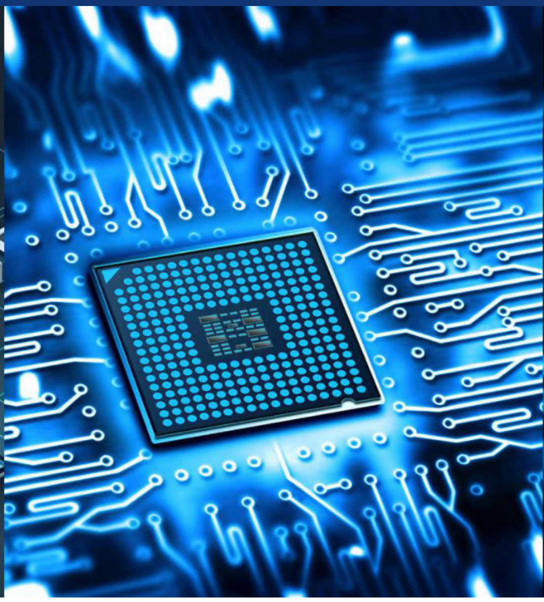
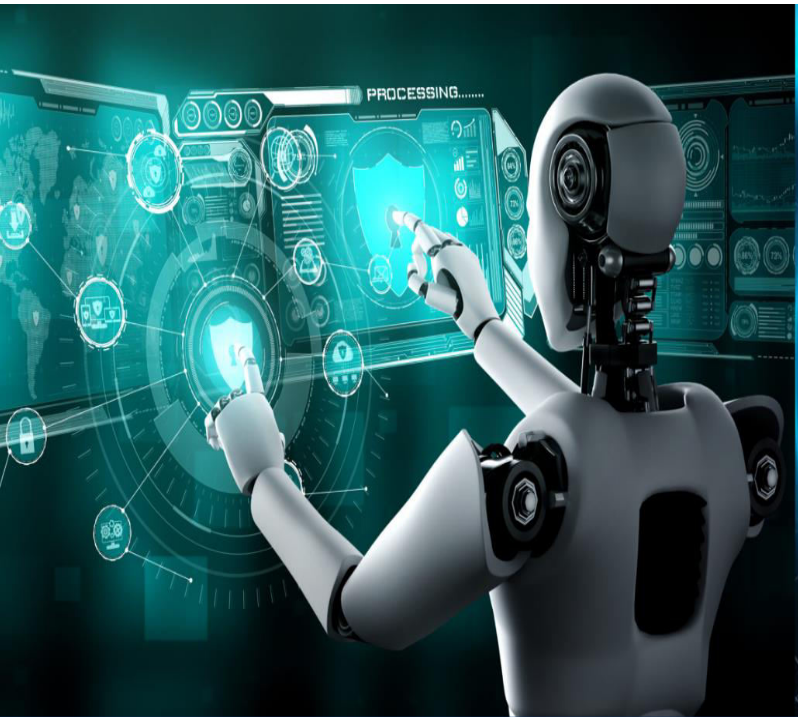
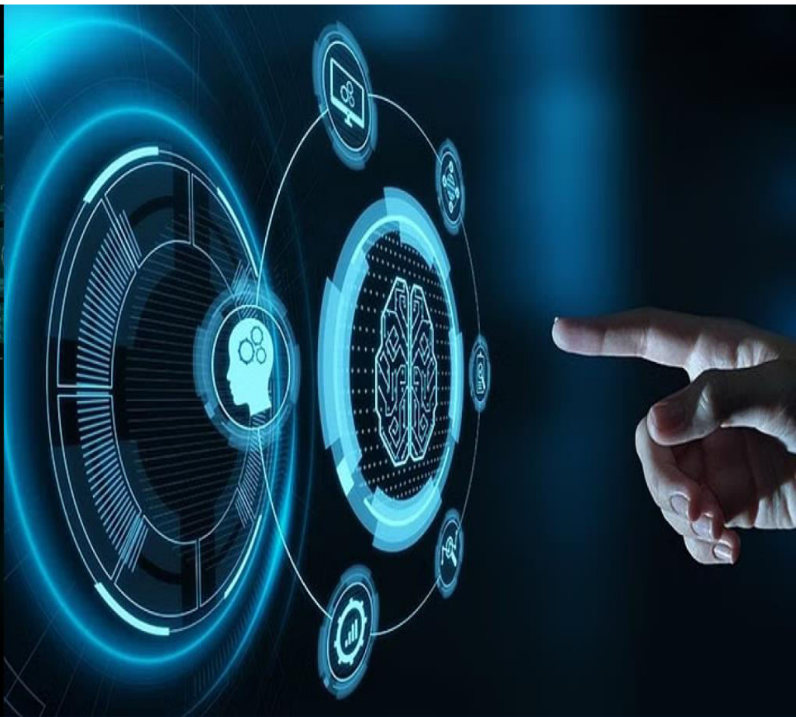
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 3, March 2025

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303007

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Person Re-Identification using Unsupervised Learning and ResNets in Surveillance Systems: A Progressive Clustering-based Learning Approach

Xuezhi Wen[1], Maambo Lubinda[2], Zamadder Md Iqbal Hossain[2], Arafat Daud Abdirahman[3]

Associate Professor, Department of CST, School of Computer Science, School of Cyber Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China[1]

M. Sc Student, Department of CST, School of Computer Science, School of Cyber Science and Engineering, Nanjing University of Information Science and Technology, Nanjing, China[2]

M. Sc Student, Department of CST, Anhui University of Science and Technology, Huainan, China[3]

ABSTRACT: Unsupervised person re-identification (re-ID) is a deep learning task aimed at extracting distinctive features without the need for labeled data, this task is crucial for practical applications in surveillance systems. Recent advancements in re-ID have shown the effectiveness of deep learning techniques in learning robust pedestrian representations. However, training deep models typically requires large-labeled datasets, which are both costly and impractical for real-world deployment. To overcome this challenge, we propose an Unsupervised Learning framework that transfers pre-trained deep representations to unseen domains with minimal reliance on labeled data. Our approach alternates between pseudo-labeling through image clustering and CNN fine-tuning, progressively refining feature representations. Additionally, we incorporate an advanced feature extraction strategy and k-means clustering for pseudo-labeling, ensuring robust unsupervised training. Extensive experiments on large-scale re-ID benchmarks, including Market-1501 and DukeMTMC-reID`, demonstrate that our method outperforms baseline models, achieving superior accuracy in both supervised and unsupervised settings. This framework offers a scalable and efficient solution for real-world person re-ID applications, effectively bridging the gap between supervised and unsupervised learning while significantly reducing annotation costs.

KEYWORDS: Person re-identification; unsupervised learning; Progressive Clustering-Based Learning; Confidence-Based Sample Selection; ResNet-50.

I. INTRODUCTION

Person re-identification (re-ID) is the system that aims to attain the re-identity of a particular person captured by different surveillance cameras. It is a fundamental task in computer vision, enabling the tracking of individuals across non-overlapping camera views in complex environments such as airports, shopping malls, and urban streets [1]. Person Re-ID has gradually come into our sight, with the advancement and development of science and technology. The earliest person Re-ID method was developed based on image data, which involve analyzing and comparing pedestrian images to determine the identity or Re-ID of individuals across different scenes. This technique has a wide range of applications in the fields of surveillance, intelligent transport, and security. However, image-based person Re-ID methods face some challenges, such as pose change, view angle change, and occlusion in a single image, which limit its accuracy and robustness in real scenes. [2]

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

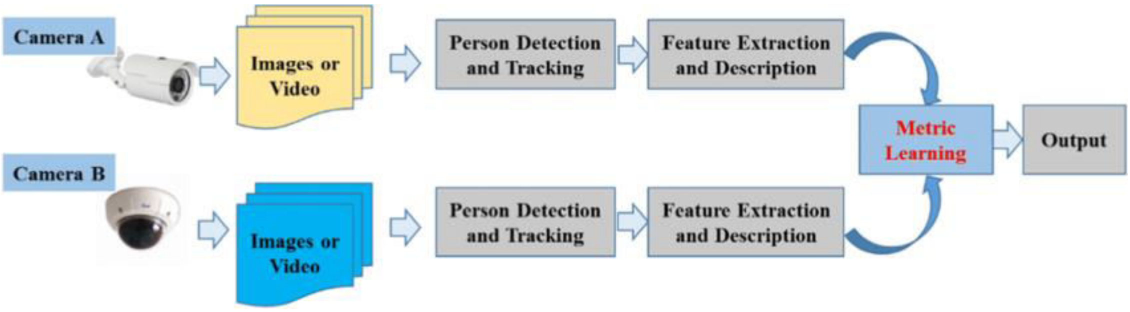


Figure 1: Flow chart illustrating the process of person re-identification, including feature extraction, clustering, and matching across camera views.

Person re-identification (re-ID) constitutes a pivotal framework engineered to recognize individuals across disparate surveillance systems, thereby serving a crucial function in applications pertaining to public safety, intelligent transportation, and security oversight. With technological advancements, re-ID methodologies have progressed from primitive image-based techniques, which depended on the analysis and comparison of pedestrian visuals, to more intricate paradigms that harness the capabilities of deep learning. Notwithstanding its extensive applications, conventional image-based re-ID techniques encounter fundamental constraints, such as variations in pose, discrepancies in viewing angles, and occlusions [3],[4], which impede their precision and resilience in practical scenarios. These challenges have catalyzed the innovation of enhanced methodologies, including deep learning architectures such as ResNets, which have markedly augmented the efficacy and scalability of re-ID systems. As the exigency for precise and efficient re-ID methodologies continues to escalate, the domain remains a vibrant sphere of inquiry, with persistent endeavors aimed at addressing its complexities and improving its practical applicability.

Common Challenges of Person Re-ID

In spite of the advancements achieved in person re-identification (re-ID) through the application of deep learning techniques, numerous challenges endure in practical implementations. A predominant concern is the inconsistency in appearance across various camera perspectives, influenced by elements such as illumination, camera orientations, and focal lengths, which can impede accurate identification. Occlusions, wherein segments of an individual are obscured, along with pose alterations, including variations in body posture, further exacerbate the complexities associated with re-ID, particularly in densely populated environments.

An additional obstacle lies in the dependence on extensive labeled-datasets for supervised learning, a process that is both time-intensive and impractical for large-scale systems that undergo frequent updates. This underscores the necessity for unsupervised learning methodologies that diminish reliance on labeled data [5].

Lastly, cross-domain discrepancies restrict the generalizability of re-ID models. Models that are trained on a specific dataset may encounter difficulties when applied to novel environments due to variations in camera technology, resolution, or environmental conditions, adversely impacting their performance when implemented in real-world contexts, such as transitioning from high-resolution to low-resolution cameras [6].

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Figure 2: Images showing the different challenges of person Re-identification.

A prominent obstacle in the domain of unsupervised re-identification (re-ID) methodologies, particularly those predicated on clustering techniques, lies in their vulnerability to erroneous clustering outcomes. The premise of clustering posits that analogous images will constitute distinct clusters for each individual; however, in practice, clustering algorithms frequently yield imprecise classifications owing to substantial variability and the presence of outliers[7]. For instance, images depicting the same individual may inadvertently be allocated to divergent clusters as a result of variations in pose or occlusions, whereas images representing different individuals may be erroneously grouped together due to similarities in attire or background. Such inaccuracies can compromise the entire training regimen, culminating in suboptimal feature learning and diminished re-ID efficacy. This underscores the imperative for the development of more resilient clustering methodologies capable of addressing noise inherent in real-world datasets.

Research Problem

The difficulties associated with person re-identification (re-ID) within genuine surveillance frameworks, particularly employing unsupervised techniques, are considerable. Although clustering-based strategies have achieved some advancements, they remain constrained by the prevalence of noisy clustering outcomes, attributable to factors such as variability in appearance, occlusions, and alterations in pose [8]. This situation engenders erroneous classifications, wherein images of the same individual may be situated in disparate clusters, or images of distinct individuals may be coalesced. These inaccuracies can propagate throughout the training continuum, thereby diminishing the quality of feature learning and re-ID performance [9].

Moreover, a substantial number of unsupervised techniques depend on proximity-based sample selection, whereby training instances are identified based on their distance from cluster centroids. While this approach is ostensibly straightforward, it neglects to account for the reliability of the selected instances, often incorporating noisy or outlier samples, which further compromises performance [10].

The reliance on extensive labeled datasets presents an additional challenge, as supervised methodologies necessitate labor-intensive and costly manual annotations, rendering them impractical for large-scale systems [4]. Although unsupervised methodologies display greater scalability, they continue to encounter difficulties in achieving performance levels comparable to those of supervised techniques, primarily due to unreliable training instances and noise within clustering.

These challenges accentuate the necessity for a more robust and adaptive unsupervised re-ID methodology capable of addressing noisy clustering outcomes and selecting dependable training instances to enhance performance.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Research Motivation

The impetus for this research is derived from the escalating demand for precise and scalable person re-identification (re-ID) systems in actual surveillance contexts, such as public safety, crime deterrence, and crowd management. Although unsupervised re-ID methods present a promising alternative to supervised techniques by obviating the requirement for labeled data, their efficacy is frequently undermined by noisy clustering results and unreliable sample selection. Mitigating these constraints is essential for the advancement of re-ID systems capable of functioning effectively within dynamic, large-scale environments. Through the proposal of Progressive Clustering-Based Learning (PCBL), this study aspires to reconcile the disparity between supervised and unsupervised re-ID performance, thereby facilitating the development of more robust and adaptable solutions to the challenges posed by real-world surveillance.

II. RELATED WORK

Existing works in the field of Re-ID are discussed in this section, which can be categorized into two main parts: (1) supervised person re-ID and image representations in latent space, and (2) unsupervised person re-ID. Supervised Person Re-ID and Image Representations in Latent Space. Deep learning models have gained significant attention since Krizhevsky et al. [11] achieved a substantial breakthrough in ILSVRC12. Recently, deep learning-based person representations have demonstrated state-of-the-art performance in person re-ID. The earliest studies incorporating deep learning into re-ID were introduced by Li et al. [12] and Ahmed et al. [13]. In general, re-ID models can be categorized into two primary architectures: classification-based models, similar to those used in conventional image classification tasks [11], and siamese network-based models, which employ image pairs [14] or triplet-based learning strategies [15]. In the early stages, when re-ID datasets were relatively small, such as VIPeR [16], which contains only two images per identity, the siamese network approach was predominantly used. However, as larger-scale datasets such as Market-1501 [41] became available, classification models gained widespread adoption within the re-ID community. A comprehensive review of re-ID methodologies can be found in [39].

From a methodological perspective, representation learning approaches have demonstrated greater scalability for handling large gallery sizes compared to deep similarity learning methods [17], [18], and [19]. For instance, Hermans et al. [20] introduced an optimized variant of the triplet loss that refines distancebased learning. Another widely adopted strategy involves training an identification network and utilizing intermediate feature representations as discriminative embeddings [39], [42], [21], [22]. Xiao et al. [23] proposed an online instance matching loss to address the challenge of limited training samples per identity, while Lin et al. [24] incorporated both attribute classification and re-ID loss to enhance feature embeddings. Additionally, Zheng et al. [40] leveraged generative adversarial networks (GANs) to generate synthetic samples, aiming to achieve a more uniform prediction distribution in the softmax layer. In this study, we adopt the baseline identification model, ID-Discriminative Embedding (IDE), introduced in [21], as the foundation for our proposed approach.

Unsupervised Person Re-ID. While supervised learning methods have led to significant advancements in deeply learned person representations, relatively less emphasis has been placed on unsupervised learning for person re-ID. Several studies have explored alternative approaches to address this challenge. Kodirov et al. [25] introduced a graph-regularized dictionary learning framework to extract discriminative identity cues for cross-view matching. Yang et al. [26] proposed a weighted linear coding approach to derive multi-level descriptors from raw image data without supervision. Wang et al. [27] leveraged a kernel subspace learning model to extract identity-specific cross-view features from unlabeled datasets. Meanwhile, Peng et al. [28] developed a multi-task dictionary learning technique that transfers a view-invariant representation from existing labeled source datasets to an unlabeled target domain. Additionally, Ma et al. [29] incorporated image sequence information to integrate multiple feature descriptors. However, these approaches have primarily focused on small-scale datasets and do not incorporate deep feature representations.

Another direction in unsupervised person re-ID involves the direct utilization of hand-crafted features. Several effective feature descriptors have been developed over the years. In earlier works, Farenzena et al. [30] proposed segmenting the pedestrian foreground from the background using weighted color histograms, maximally stable color regions, and recurrent high-structured patches. Gray and Tao [16] introduced a method that applies 8 color channels and 21 texture filters on the luminance channel while partitioning pedestrian images into horizontal stripes. More recent studies by

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Zhao et al. [31], [32], [33] employed a 32-dimensional LAB color histogram and a 128-dimensional SIFT descriptor, extracting these features from densely sampled 10×10 pixel patches with a step size of 5 pixels. Liao et al. [34] proposed the Local Maximal Occurrence (LOMO) descriptor, integrating color and SILTP histograms, which was later utilized by subsequent works [35], [36]. Similarly, Chen et al. [37] adopted a related set of handcrafted features. Zheng et al. [41] introduced an 11-dimensional color names descriptor, extracting local patch-level features and aggregating them into a global representation using a Bag-of-Words model.

III. METHODOLOGY

To address this problem, we have developed an unsupervised training approach to give the knowledge of recently viewed images by the camera to the model without manual labelling. We have two main parts in this experiment, which are supervised training (Base model initialization) and unsupervised training (progressive learning). The following subsections are the detailed explanation of our proposed method.

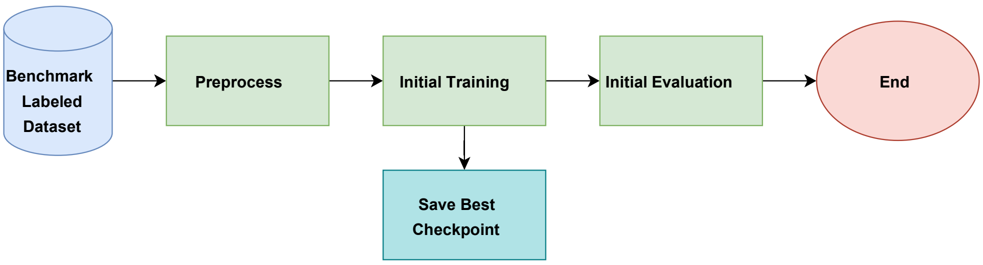


Figure 3: Supervised Training on labeled dataset.

Our model operates in a progressive manner, starting with a small set of reliable samples and gradually expanding to include more as the model improves. Initially, the framework focuses on samples located near cluster centroids, which are more likely to be correctly labeled. As the model becomes more confident in its predictions, it adaptively incorporates a broader range of samples into the training process. This self-paced learning mechanism ensures steady and robust improvements in feature learning.

The framework begins with a pre-trained CNN model (ResNet-50) fine-tuned on a Benchmark dataset called Market1501 [41]. This is the supervised part of our method, where we initialize the base model for the unsupervised training. Figure 1 shows the flowchart of the supervised training phase. We preprocess the dataset, such as resizing, padding, transformations, and augmentations. We have normalized the pixel values for more numeric stability. During the training, we have used an adaptive learning rate using a learning rate scheduler for optimal convergence. In our training, we have used Stochastic Gradient Descent (SGD) optimizer and Cross-Entropy as the loss function. We trained the supervised model for 60 epochs and took the best checkpoint as the base model for unsupervised training. In the model architecture, we added a dense classifier layer to get the ID of the input person. Our main purpose here is to learn the embedding of the person's image while doing the classification task so that we can rank the image from the gallery (captured using the cameras) to determine if the person is reappearing or not. After the training phase, we saved the best checkpoint and evaluated the model. This checkpoint is used to fine-tune, using new unlabeled images while in the progressive learning phase.

i. Feature Extraction

For extracting the image feature, we have concatenated some additional information along with the feature vector from the fine-tuned model. At first, we apply horizontal flip to the image and extract the feature vector from it using the fine-tuned ResNet-50. Then, we transformed the image with some random zooming in and zooming out and extracted the feature vector of it. At last, we extract the feature vector of the original image and concatenate the three feature vectors to represent the final feature vector of the image

The feature vector x_i for each image is computed as:

$$f_i = \phi(xi; \theta_t) \forall xi \in X \quad \text{eq. 1}$$

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

ii.Progressive Clustering Unsupervised Learning

For extracting the image feature, we have concatenated some additional information along with the feature vector from the fine-tuned model. At first, we apply horizontal flip to the image and extract the feature vector from it using the fine-tuned ResNet-50. Then, we transformed the image with some random zooming in and zooming out and extracted the feature vector of it. At last, we extract the feature vector of the original image and concatenate the three feature vectors to represent the final feature vector of the image.

Progressive Unsupervised Learning

Figure 1 shows the flowchart of the unsupervised training phase. We have taken another benchmark dataset named DukeMTMC-reID [5] and used it as the unlabeled dataset. We have preprocessed the dataset in the same manner that we previously discussed in section 5.1. A pre-defined limit was set as the number of iterations. We have also monitored the convergence of the loss during the training iterations. If we did not reach the iteration limit, then fine tune the base model with the unlabeled data. For doing this training, we need to perform some pre-step processes.

In the first iteration, we used the base model for extracting the feature vector from each image. From the second iteration, we used the previous iteration's best checkpoint for extracting the feature vector. After extracting the features, we apply the K-Means algorithm to cluster the unlabeled images. We use the cluster number as the class label for the images inside the same cluster.

The clustering process is defined as:

$$\{c_1, c_2, \dots, c_k\} = k - \mathrm{means}\left(\{f_{1,} f_{2,\dots,} f_N\}, K\right) \qquad \text{eq. 2}$$

where K is the number of clusters (equal to the number of identities), and c_k is the centroid of the k-th cluster. Each sample x_i is assigned to a cluster based on its proximity to the cluster centroid.

iii. Confidence-Based Sample Selection

To reduce the impact of noisy clustering results, we introduce a confidence-based sample selection mechanism. We have filtered out some images based on the cluster confidence for fine-tuning the base model. We have selected the training sample where the confidence is more than 85%. This confidence threshold is pre-defined and configurable. After we add the pseudo label to the images, we fine-tune the base model in the same manner that we demonstrated in figure 3. We used the best checkpoint of this iteration to extract the feature vector in the next iteration. Since this checkpoint has the knowledge of the unlabeled images, in the next iteration, when we use the feature vectors from this model for clustering the images, it will be more precise and accurate. Thus, gradually, the model will learn and improve its performance. We ran 10 iterations in this experiment and selected the best checkpoint to evaluate our proposed method.

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303007

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

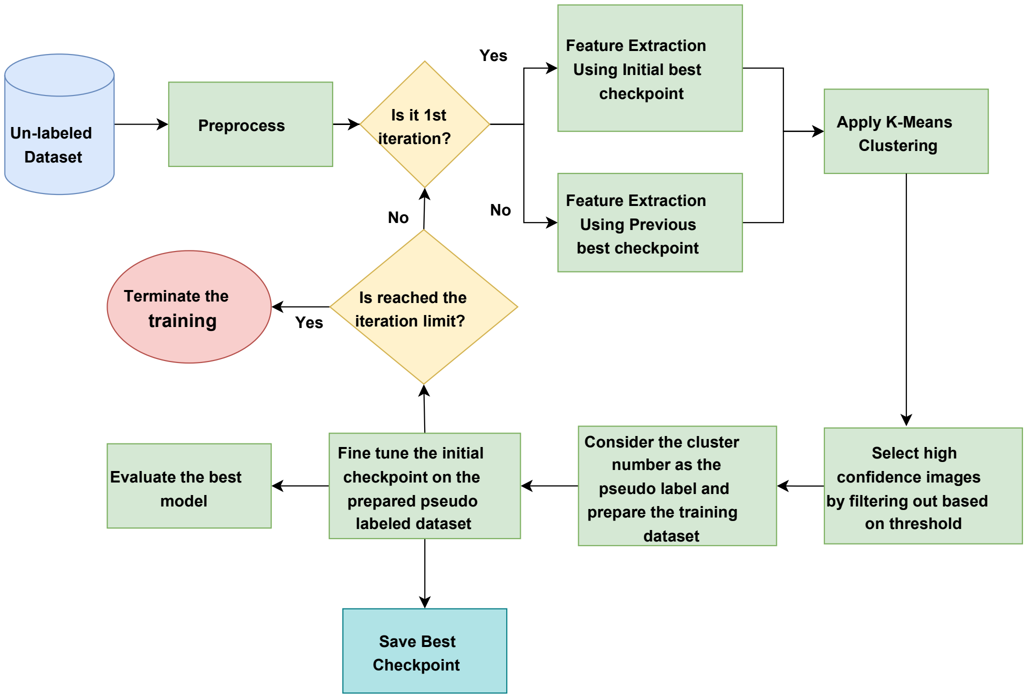


Figure 4: Unsupervised Training on Unlabelled Images.

IV. EVALUATION METRICS AND EXPERIMENTAL RESULTS

We evaluate our method using Rank-1 accuracy, which measures the percentage of correct matches in the top result, and mean Average Precision (mAP), which assesses overall retrieval performance by averaging precision across recall levels. These metrics are widely used in person re-identification (Re-ID) to evaluate both accuracy and robustness.

- Rank-1 Accuracy: This metric measures the percentage of times the correct identity appears in the top 1 ranked retrieval result. It reflects the model's ability to precisely identify the true match at the highest rank, which is critical for real-world applications where the top result is often the most important.
- Mean Average Precision (mAP): mAP evaluates the overall retrieval performance by averaging precision at each recall level. It considers the model's ability to rank all correct matches higher while minimizing false positives, providing a comprehensive measure of retrieval quality across the entire ranking list.

Together, these metrics offer a balanced evaluation of Re-ID models, capturing both the precision of top-ranked results and the consistency of retrieval performance across all relevant matches.

We outline the Rank-1 accuracy and mean average precision (mAP) measurements for each dataset to affirm the efficiency of our suggested method. These findings furnish a thorough assessment of the performance of the PCBL approach in relation to the prevailing state-of-the-art Re-ID methodologies.

We have trained two base models to evaluate our method. We have selected some baseline models from existing works for the comparison. Table 1 shows that, the proposed model outperformed the baseline models in supervised training on DukeMTMC-reID dataset. We have trained the model using the training dataset of DukeMTMC-reID and evaluated it using the test dataset.

We compare our proposed model to state-of-the-art unsupervised domain adaptation methods and purely unsupervised methods for person Re-ID. The purely unsupervised methods for person Re-ID include: PUL (baseline), LOMO, BOW, and UMDL. The results are summarized in the table below, highlighting the performance of our method across Market-

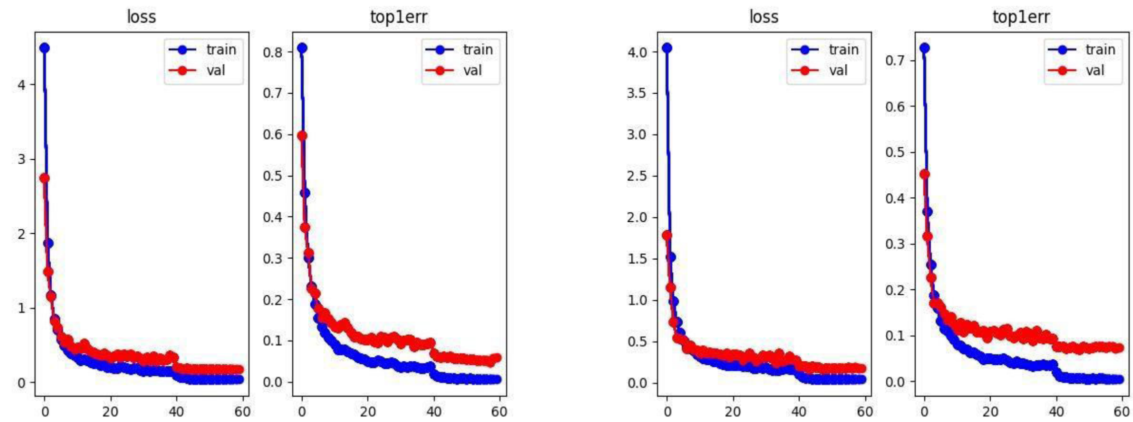
© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303007

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

1501, DukeMTMC-reID, and CUHK03 datasets. Our proposed framework demonstrates consistent improvements over the compared method, achieving higher Rank-1 accuracy and mAP across all datasets. For instance, it achieved 80.386% for rank-1, 90.395% for rank-5, 93.626% for rank-10 accuracy, and mAP of 65.333%. This result outperformed the baseline model for all the metrics. Table 2 shows the supervised training result on Market1501 dataset. We observed that the proposed model outperformed the baseline models in every metric by achieving 80.3860% for rank-1, 90.3950% for rank-5, 93.6266% for rank-10 accuracy, and mAP of 65.3330%. The figure 3 shows the loss and top 1 error convergence during the validation phase. We can observed that, within 60 epochs, it has reached the plateau point.



(a) On Market1501 Dataset

(b) On DukeMTMC-reID Dataset

Figure 5: Loss and Top 1 error of training and validation dataset during supervised training on Market1501 and DukeMTMC-reID.

These results underscore the effectiveness of our approach in improving both precision and robustness in unsupervised person Re-ID tasks.
The findings of this assessment are concisely presented in the table below:

Table 1: Performance comparison table of our proposed model and other state of the art models:

Method	Market1501		DukeMTMC-reID	
	Rank 1	mAP	Rank 1	mAP
CAMEL [40]	54.5	26.3	-	-
PUL [7]	45.5	20.5	30.0	16.4
PTGAN [41]	58.1	26.9	46.9	26.4
TJ-AIDL [42]	58.2	26.5	44.3	23.0
HHL [43]	62.2	31.4	46.9	27.2
MAR [44]	67.7	40.0	67.1	48.0
ENC [45]	75.1	43.0	63.3	40.4
ATNet [46]	55.7	25.6	45.1	24.9
PAUL [47]	68.5	40.1	72.0	53.2
SBGAN [48]	58.5	27.3	53.5	30.8
UCDA [49]	64.3	34.5	55.4	36.7
CASC [50]	65.4	35.5	59.3	37.8

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

PDA [51]	75.2	47.6	63.2	45.1
CR-GAN [52]	77.7	54.0	68.9	48.6
PAST [53]	78.4	54.6	72.4	54.3
SSG [54]	80.0	58.3	73.0	53.4
HCT [55]	80.0	56.4	69.6	50.7
SNR [56]	82.8	61.7	76.3	58.1
MMT [57]	87.7	71.2	78.0	65.1
MEB-Net [58]	89.9	76.0	79.6	66.1
BoW [4]	9.04	3.26	-	-
BoW+Geo [4]	21.23	8.46	-	-
BoW+Geo+Gauss [4]	34.38	14.10	-	-
BoW+Geo+Gauss+MQ [4]	41.21	17.63	-	-
MaskR-CNN [59]	71.64	65.54	-	-
Mask R-CNN + GrabCut [59]	84.51	80.98	-	-
Proposed Method	**89.9347**	**74.8229**	**80.3860**	**65.3330**

The following graphs illustrate the Rank-1 accuracy of different methods across three datasets: Market-1501, DukeMTMC-reID, and CUHK03. These comparisons highlight the performance of the baseline model (PUL) and the proposed model (PCBL).

V. CONCLUSION AND FUTURE WORK

This paper introduces Progressive Clustering-Based Learning (PCBL), which leverages an iterative process between k-means clustering and CNN finetuning. Our findings highlight that we can identify new images from multiple camera sources, without manual annotations, using the proposed method. Utilizing the ResNet-50 architecture, our model enhances the extraction of discriminative feature representations through deep residual learning, which significantly improves the robustness and reliability of unsupervised person Re-ID. By selecting reliable training samples based on model confidence, the model progressively enhances feature learning, demonstrating superior adaptability across diverse camera views and varying surveillance environments. These improvements enable the model to serve as a more effective and scalable solution for real-world applications, overcoming some limitations in similar prior methods.

The proposed approach is simple to implement and offers potential for further enhancements. For instance, in video-based re-ID, frames within a tracklet can be assumed to belong to the same identity, which can serve as an initialization strategy for clustering and fine-tuning. Additionally, incorporating diversity into PUL presents another promising direction, enabling the selection of training samples across multiple camera views to further improve fine-tuning.

For future work, it is essential to explore the integration of cross-camera diversity in sample selection, ensuring that training data encompasses a variety of camera perspectives to enhance generalization. Moreover, while k-means clustering is widely used, it may encounter difficulties in handling outliers or irregular cluster shapes. The adoption of more advanced clustering techniques such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) or spectral clustering could increase the robustness of the clustering process, particularly in the presence of complex data distributions or noise. Furthermore, the integration of deep clustering methods, which jointly learn the clustering structure and feature representations, could significantly enhance overall accuracy and scalability, presenting promising avenues for further research in improving unsupervised person Re-ID systems.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1.Mallika Alias Pandeeswari, R., & Rajakumar, G. (2025). Deep intelligent technique for person Re-identification system in surveillance images.Pattern Recognition, 162 111349. (https://doi.org/10.1016/j.patcog.2025.111349)

2.Wang, J., Gao, X., Zhu, F., & Chen, X. (2024). Exploring frontier technologies in video-based person re-identification: A survey on deep learning approach. Pattern Recognition, 162, 111349. (https://doi.org/10.1016/j.patcog.2024.111349)

3.Qian, Y., Barthelemy, J., Karuppiah, E., et al. (2024). Identifying Re-identification Challenges: Past, Current and Future Trends. SN COMPUT. SCI., 5, 937. (https://doi.org/10.1007/s42979-024-03271-9)

4.Zahra, A., Perwaiz, N., Shahzad, M., & Fraz, M. M. (2023). Person re-identification: A retrospective on domain specific open challenges and future trends. Pattern Recognition, 142, 109669. (https://doi.org/10.1016/j.patcog.2023.109669)

5.Klingler, N. (2023). Deep Learning for Person Re-Identification. VISO.AI. (https://viso.ai/deep-learning/deep-learning-for-person-re-identification/)

6.Han, C., Jiang, B., & Tang, J. (2022). Multi-granularity cross attention network for person re-identification. Multimedia Tools and Applications, 82. (https://doi.org/10.1007/s11042-022-13833-9)

7.X. Zhu, Y. Li, J. Sun, H.-J. Chen, and J. Zhu, "Learning with Noisy Labels Method for Unsupervised Domain Adaptive Person Re-identification," Neurocomputing, vol. 452, pp. 1-12, May 2021.(doi: 10.1016/j.neucom.2021.04.120)

8.Zhang, X., Luo, H., Fan, X., Xiang, W., & Sun, Y. (2021). AlignedReID: Surpassing human level performance in person reidentification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(8), 12341246. (https://ieeexplore.ieee.org/document/9356346)

9.Fan, H., Zheng, L., & Yang, Y. (2022). Unsupervised person re-identification: Clustering and fine-tuning. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(4), 1–18. (https://doi.org/10.1145/3243316)

10.Li, A., Liu, L., & Yan, S. (2014). Person re-identification by attribute-assisted clothes appearance. In Person Re-Identification (pp. 119–138). Springer. (https://doi.org/10.1007/978-1-4471-6296-4_6)

11.Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105. (https://doi.org/10.1145/3065386)

12.Yi, D., Lei, Z., Liao, S., & Li, S. Z. (2014). Deep metric learning for person re-identification. In Proceedings of the 22nd International Conference on Pattern Recognition (pp. 34–39). IEEE. (https://doi.org/10.1109/ICPR.2014.16)

13.Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). DeepReID: Deep filter pairing neural network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 152–159). (https://doi.org/10.1109/CVPR.2014.27)

14.Radenović, F., Tolias, G., & Chum, O. (2016). CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples. In Computer Vision–ECCV 2016: 14th European Conference (Vol. 9910, pp. 3–20). Springer. (https://doi.org/10.1007/978-3-319-46466-4_1)

15.Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 815–823). (https://doi.org/10.1109/CVPR.2015.7298682)

16.Gray, D., & Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In Computer Vision–ECCV 2008: 10th European Conference (Vol. 5303, pp. 262–275). Springer. (https://doi.org/10.1007/978-3-540-88682-2_21)

17.Varior, R. R., Haloi, M., & Wang, G. (2016). Gated siamese convolutional neural network architecture for human re-identification. In Computer Vision–ECCV 2016: 14th European Conference (Vol. 9910, pp. 791–808). Springer. (https://doi.org/10.1007/978-3-319-46466-4_49)

18.Liu, H., Feng, J., Qi, M., Jiang, J., & Yan, S. (2017). End-to-end comparative attention networks for person re-identification. IEEE Transactions on Image Processing, 26(7), 3492–3506. (https://doi.org/10.1109/TIP.2017.2692268)

19.Chen, W., Chen, X., Zhang, J., & Huang, K. (2017). Beyond triplet loss: A deep quadruplet network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 403–412). (https://doi.org/10.1109/CVPR.2017.145)

20.Hermans, A., Beyer, L., & Leibe, B. (2017). In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737. (https://doi.org/10.48550/arXiv.1703.07737)

21.Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., & Tian, Q. (2016). MARS: A video benchmark for large-scale person re-identification. In Computer Vision–ECCV 2016: 14th European Conference (Vol. 9910, pp. 868–884). Springer. (https://doi.org/10.1007/978-3-319-46466-4_53)

22.Zheng, L., Huang, Y., Lu, H., & Yang, Y. (2019). Pose-invariant embedding for deep person re-identification. IEEE Transactions on Image Processing, 28(9), 4500–4509. (https://doi.org/10.1109/TIP.2019.2915299)

23.Xiao, T., Li, H., Ouyang, W., & Wang, X. (2017). Joint detection and identification feature learning for person search. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3415–3424). (https://doi.org/10.1109/CVPR.2017.406)

24.Lin, Y., Zheng, L., Zheng, Z., Wu, Y., Hu, Z., Yan, C., & Yang, Y. (2019). Improving person re-identification by attribute and identity learning. Pattern Recognition, 95, 151–161. (https://doi.org/10.1016/j.patcog.2019.06.017)

25.Kodirov, E., Xiang, T., Fu, Z., & Gong, S. (2016). Person re-identification by unsupervised ℓ_1 graph learning. In European Conference on Computer Vision (pp. 1–16). Springer. (https://doi.org/10.1007/978-3-319-46493-0_25)

26.Yang, Y., Wen, L., Lyu, S., & Li, S. (2017). Unsupervised learning of multilevel descriptors for person re-identification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1). (https://doi.org/10.1609/aaai.v31i1.11223)

27.Wang, H., Zhu, X., Xiang, T., & Gong, S. (2016). Towards unsupervised open-set person re-identification. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 769–773). IEEE. (https://doi.org/10.1109/ICIP.2016.7532499)

28.Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T., & Tian, Y. (2016). Unsupervised cross-dataset transfer learning for person re-identification. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1306–1315). IEEE. (https://doi.org/10.1109/CVPR.2016.146)

29.Ma, X., Zhu, X., Gong, S., Xie, X., Hu, J., Lam, K.-M., & Zhong, Y. (2017). Person re-identification by unsupervised video matching. Pattern Recognition, 65, 197–210. (https://doi.org/10.1016/j.patcog.2017.01.001)

30.Farenzena, M., Bazzani, L., Perina, A., Murino, V., & Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. In 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (pp. 2360–2367). IEEE. (https://doi.org/10.1109/CVPR.2010.5539967)

31.Zhao, R., Ouyang, W., & Wang, X. (2013). Unsupervised salience learning for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3586–3593). (https://doi.org/10.1109/CVPR.2013.459)

32.Zhao, R., Ouyang, W., & Wang, X. (2013). Person re-identification by salience matching. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2528–2535). https://doi.org/10.1109/ICCV.2013.318

33.Zhao, R., Ouyang, W., & Wang, X. (2014). Learning mid-level filters for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 144–151). (https://doi.org/10.1109/CVPR.2014.27)

34.Liao, S., Hu, Y., Zhu, X., & Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2197–2206). (https://doi.org/10.1109/CVPR.2015.7298764)

35.Zhang, L., Xiang, T., & Gong, S. (2016). Learning a discriminative null space for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1239–1248). (https://doi.org/10.1109/CVPR.2016.143)

36.Zhang, Y., Li, B., Lu, H., Irie, A., & Ruan, X. (2016). Sample-specific SVM learning for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1278–1287). (https://doi.org/10.1109/CVPR.2016.141)

37.Chen, D., Yuan, Z., Chen, B., & Zheng, N. (2016). Similarity learning with spatial constraints for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1268–1277). (https://doi.org/10.1109/CVPR.2016.141)

38.He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770–778). (https://doi.org/10.1109/CVPR.2016.90)

39.Zheng, L., Yang, Y., & Hauptmann, A. G. (2016). Person re-identification: Past, present and future. arXiv preprint arXiv:1610.02984. (https://doi.org/10.48550/arXiv.1610.02984)

40.Zheng, Z., Zheng, L., & Yang, Y. (2017). Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3774–3782). (https://doi.org/10.1109/ICCV.2017.405)

41.Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1116–1124). https://doi.org/10.1109/ICCV.2015.134

42. Xiao, T., Li, H., Ouyang, W., & Wang, X. (2016). Learning deep feature representations with domain guided dropout for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1249–1258). https://doi.org/10.1109/CVPR.2016.155