DOI 10.15622/ia.24.3.9

M. IDRISSI ALAMI, A. EZ-ZAHOUT, F. OMARY

COMPARATIVE STUDY OF PERSON RE-IDENTIFICATION TECHNIQUES BASED ON DEEP LEARNING MODELS

Idrissi Alami M., Ez-zahout A., Omary F. **Comparative Study of Person Re-Identification Techniques Based on Deep Learning Models.**

Abstract. Person re-identification (Re-ID) is crucial in intelligent surveillance, requiring precise identification of individuals across multiple camera viewpoints. Traditional distance-based methods, such as Euclidean and Cosine, struggle with challenges like posture variations and occlusions, limiting their effectiveness. This study explores deep metric learning models, specifically Siamese and Triplet networks, to improve Re-ID performance. We evaluate these methods on the Market-1501 dataset using Cumulative Matching Characteristic (CMC) and Cumulative Distribution Function (CDF) curves. Our findings reveal that the Triplet network outperforms traditional approaches at higher ranks, achieving Rank-5 accuracy of 78.6% and Rank-10 accuracy of 93%, while its Rank-1 accuracy remains low (0.06%). In contrast, Euclidean and Cosine distances show poor Rank-1 performance (2% and 0.30%, respectively), highlighting their limitations. Additionally, incorporating VGG16 enhances feature extraction, improving recognition by capturing fine-grained spatial details. This comparative study highlights the effectiveness of deep metric learning and underscores its potential for real-world surveillance applications. However, the computational demands of deep networks present challenges for real-time deployment. Future research should focus on optimizing model efficiency, reducing computational costs, and extending evaluations to real-time scenarios.

Keywords: CMC/CDF metrics, convolutional neural networks (CNNs), deep learning, person re-identification, VGG16, video surveillance systems.

1. Introduction. Video surveillance systems progress through four fundamental stages: Detection, Tracking, Profile Analysis, and Re-Identification (Re-ID). Detection acts as the base by recognizing humans or things inside a monitored region, enabling additional analysis and tracking [1]. Tracking tracks subjects across scenes or cameras, maintaining continuity. Profile Analysis captures crucial aspects such as looks and behavior for identification. Re-ID links humans across sites, which is important for seamless surveillance. Together, these strategies increase accuracy, especially in busy settings and complicated security setups [2].

Person re-identification (re-ID) is a crucial problem in video surveillance systems, concentrating on identifying individuals across different camera viewpoints. The ability to precisely re-identify someone has important ramifications for security, forensic investigations, and monitoring applications. Despite tremendous development in this sector, achieving high accuracy remains a problem due to changes in appearance, stance, lighting conditions, and occlusions [3].

Early person re-identification algorithms generally relied on handcrafted feature representations and simple distance metrics such as Euclidean distance, cosine similarity, and Mahalanobis distance. The

Euclidean distance, one of the simplest metrics, calculates the straight-line distance between feature vectors, making it computationally efficient but ineffectual for complicated variations in surveillance scenarios [4]. Similarly, cosine similarity is widely used for high-dimensional feature spaces, measuring the angular difference between feature vectors rather than absolute distances. However, it struggles to provide correct matches at Rank-1 and exhibits reduced accuracy at Rank-5 [5]. The Mahalanobis distance contains covariance information to account for correlations between features, exhibiting greater performance in some scenarios, such as fluctuating lighting and background clutter [6, 7]. However, these methods alone often fail to generalize well across different datasets and environmental conditions.

To circumvent these constraints, deep learning algorithms, particularly Convolutional Neural Networks (CNNs), have been widely adopted due to their capacity to learn hierarchical and discriminative feature representations directly from images [5, 8]. Prominent CNN architectures, including VGG16, ResNet, and Inception, have significantly boosted person re-identification performance. Specifically, VGG16 has exhibited exceptional results in image classification, transfer learning, and capturing detailed spatial characteristics beneficial for re-identification [6, 9, 10, 11]. However, its deep structure demands substantial computational resources, prompting the investigation of more efficient architectures such as ResNet and Inception [7, 12]. Additionally, deep learning advancements have fostered the development of metric learning frameworks like Siamese and Triplet networks, which have proven effective in distinguishing individuals with similar appearances across different camera perspectives [8].

Beyond standard CNNs, metric learning-based techniques such as Siamese networks and Triplet networks have significantly boosted re-identification performance. Siamese networks leverage contrastive loss to learn a feature space where related identities are clustered together while distinct identities are pushed apart [13]. While they outperform traditional distance metrics at higher ranks (e.g., Rank-5), they still struggle at Rank-1 [14]. Triplet networks improve this principle by including anchor-positive-negative triplets, guaranteeing that positive samples are closer to the anchor than negative samples. This strategy considerably improves accuracy at Rank-5 and Rank-10 but involves complex training procedures and large-scale labeled data [15].

For evaluation, the Market-1501 dataset has become a typical benchmark for person re-identification, presenting realistic surveillance scenarios with problems such as illumination fluctuations, occlusions, and backdrop complexity [15, 16]. Performance assessment often depends on metrics such as the Cumulative Matching Characteristic (CMC) and the Cumulative Distribution Function (CDF), which provide rank-based and distributional insights into algorithm efficacy [17].

Despite breakthroughs in re-identification, problems exist, including domain adaptability, occlusions, computing limits, and privacy issues [18]. This study rigorously assesses various re-identification techniques – Euclidean, Cosine, Mahalanobis, Siamese, and Triplet networks – across standard datasets and rank-based criteria. By assessing their strengths and limits in diverse surveillance contexts, we propose modifications to deep learning-based techniques to enhance accuracy while maintaining efficiency.

In recent years, state-of-the-art person re-identification models particularly those based on Vision Transformers and hybrid CNN-ViT [15, 22] architectures have achieved over 95% Rank-1 accuracy on the Market-1501 benchmark. Despite these advancements, many of these models demand substantial computational resources, limiting their applicability in lightweight or real-time systems. In contrast, our study investigates a Triplet-based model leveraging VGG16, which, while achieving modest Rank-1 accuracy, demonstrates competitive performance at higher ranks (Rank-5 and Rank-10), making it a candidate for scalable and efficient re-identification pipelines (Table 1).

Table 1. Comparative Performance of Resent Person Re-ID Models on Market-1501

Model	Dataset	Rank-1	Rank-5	Rank-10	Notes
Our Triplet Net 2025	Market-1501	0.06%	78.6%	93%	VGG16 based
Squeeze-Net + DAE 2021 [20]	Market-1501	86.2%	92.4%	96.1%	Hybred CNN
TransReID 2023 [21]	Market-1501	95.2%	-	-	ViT-based
CNN-ViT-Loss 2025 [15]	Market-1501	93.5%	96.2%	97.4%	ViT and Loss Based
UntransReID 2024 [22]	Market-1501	95.7	-	-	Transformer based

Our study demonstrates that Triplet networks significantly outperform traditional distance metrics, achieving Rank-10 accuracy of 93%, thus contributing valuable insights into the practical deployment of intelligent surveillance systems.

The findings contribute to the continuing discourse on intelligent video surveillance, suggesting effective ways for robust and scalable human re-identification.

2. Method

2.1. Data Source. The evaluation in this work leverages the widely famous Market-1501 dataset as the core dataset for individual re-identification. Market-1501 is a benchmark dataset specifically created for person re-identification tasks in video surveillance systems. It consists of high-resolution photos acquired from six cameras in an outside location, simulating real-world problems such as differences in lighting conditions, occlusions, and varied views [15, 16].

Each individual in the collection appears in several photographs under varied situations, offering a diversified and demanding re-identification job. The dataset is separated into training and testing sets, with ground truth annotations provided, including identification labels and bounding boxes. These labels assist performance evaluation using common metrics such as Cumulative Matching Characteristics (CMC) and Cumulative Distribution Function (CDF).

2.2. Preprocessing and Feature Extraction. Before training our deep learning models, rigorous preprocessing steps are applied to ensure consistency and maximize model performance. Initially, all raw input images are resized uniformly to dimensions of 224×224 pixels, compatible with the input requirements of deep CNN architectures. To ensure effective training, pixel values are then scaled through normalization to fall within the range [0, 1], significantly enhancing the convergence stability and effectiveness of the training process. Moreover, we apply various data augmentation strategies – including random cropping, horizontal flipping, and brightness adjustments – to simulate real-world variations and further improve the robustness and generalization capabilities of our models.

For feature extraction, we leverage the well-established VGG16 architecture, pre-trained on the extensive ImageNet dataset [23]. This choice is justified by VGG16's demonstrated capability to capture rich hierarchical representations, effectively encoding discriminative characteristics such as clothing patterns, textures, and body shapes, crucial for distinguishing individuals. Specifically, we extract high-level feature vectors from the fully connected (FC) layer preceding the final softmax layer, thereby obtaining compact 128-dimensional embeddings suitable for distance-based metric learning.

These extracted embeddings are then systematically integrated into our deep learning-based re-identification frameworks, specifically Siamese and Triplet Networks. The Siamese network uses these embeddings to construct query-gallery image pairs, optimizing similarity predictions via binary cross-entropy loss, while the Triplet network processes image triplets (anchor-positive-negative), optimizing relative distances through the triplet loss. Both networks utilize the Adam optimizer with a learning rate of 0.0001, batch normalization, dropout regularization, and a batch size of 16. The learned representations are ultimately evaluated using Euclidean distance metrics, with results visualized through Cumulative Matching Characteristic (CMC) curves and Cumulative Distribution Function (CDF) plots.

2.3. Distance Metrics. In the realm of people re-identification, the choice of distance metric is pivotal for accurately quantifying the similarity between feature vectors. We evaluate several widely used metrics in this context:

– **Euclidean distance** between two feature vectors, **X** and **Y**, is computed as follows:

$$D_{Euclidean}(X, Y) = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2}, \tag{1}$$

– **Cosine distance** measures the cosine of the angle **θ** between two feature vectors, X and Y, in a high-dimensional space:

$$D_{Cosine}(X, Y) = \frac{\sum_{i=1}^{n} X_i . Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2} \cdot \sqrt{\sum_{i=1}^{n} Y_i^2}}, \tag{2}$$

– **Mahalanobis distance** between feature vectors X and Y is defined as:

$$D_{Mahalanobis}(X, Y) = \sqrt{(X - Y)^T . C^{-1} . (X - Y)}, \tag{3}$$

where C is the covariance matrix.

– **Siamese networks** utilize a neural network architecture to learn a similarity metric. The loss function for Siamese networks can be expressed as:

$$L_{Siamese}(X,Y,S) = \frac{1}{2}(1-S).D_{Euclidean}(X,Y)^2 + \frac{1}{2}S.max(0, m - D_{Euclidean}(X,Y)^2, \quad (4)$$

where **S** is the binary similarity label, and **m** is the margin.

– **Triplet networks** optimize the embedding space by minimizing the distance between similar images and maximizing the distance between dissimilar ones. The triplet loss function is defined as:

$$L_{Triplet}(A,P,N) = max(0, D_{Euclidean}(A,P) - D_{Euclidean}(A,N) + \alpha), \quad (5)$$

where A, P, and N represent anchor, positive, and negative samples, respectively, and **α** is the margin.

These distance metrics serve as fundamental components in our evaluation, contributing to the robustness and effectiveness of our intelligent video surveillance system.

2.4. Deep Learning Models and Training. We integrate the VGG16 architecture to extract high-level features from person images. The deep learning model enhances the system's ability to capture intricate patterns and representations, contributing to improved re-identification accuracy [24].

We train deep learning-based person re-identification models using both Siamese and Triplet Networks, each optimizing feature similarity differently:

The Siamese Network is designed to compare pairs of images by concatenating the feature vectors of a query and gallery image [25]. This concatenated representation is passed through fully connected layers with 512 and 256 neurons, followed by a sigmoid activation function to predict similarity. The network is trained using binary cross-entropy loss and optimized with Adam at a learning rate of 0.0001. To improve generalization and prevent overfitting, batch normalization and dropout are applied. Training is conducted with a batch size of 16, and a data generator is employed to efficiently construct query-gallery image pairs. Performance is measured by Euclidean distance calculations and Cumulative Matching Characteristic (CMC) curves.

The Triplet Network expands the learning paradigm by processing triplets consisting of an anchor, a positive, and a negative sample [26]. The model is trained using triplet loss, guaranteeing that the distance between an anchor and a positive sample (same identity) is minimized, while the distance between an anchor and a negative sample (different identity) is maximized. The architecture consists of fully connected layers with 512 and 256 neurons, followed by a sigmoid activation function. Similar to the Siamese Network, Adam optimizer is used, with a batch size of 16. The learned embeddings are evaluated using pairwise Euclidean distance metrics, and performance is visualized through CMC curves and Cumulative Distribution Function (CDF) plots.

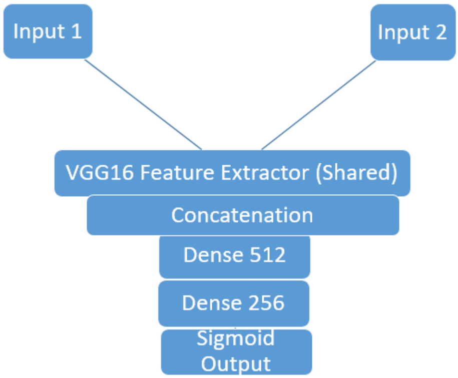

Fig. 1. Deep Learning Model based on Siamese Network

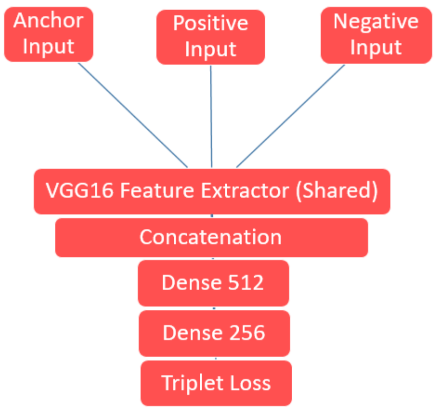

Fig. 2. Deep Learning Model based on Triplet Network

2.5. Evaluation Metrics. We assess the performance using Cumulative Matching Characteristics (CMC) and Cumulative Distribution Function (CDF) metrics. CMC illustrates the probability of correct identification within the top-k ranked matches, while CDF provides a cumulative distribution of matching scores [17, 27].

– **Cumulative Matching Characteristics (CMC):** Measures the probability of correctly identifying a person within the top-k matches, the ***CMC*** curve is formulated as follows:

$$CMC(k) = \frac{NumCIR{\leq}k}{TNQ}, \tag{6}$$

where: *NumCIR* – number of correct identifications at rank, *TNQ* – total number of queries, *k* – the rank of the match.

– **Cumulative Distribution Function (CDF):** Provides a statistical overview of matching score distributions, the CDF curve is formulated as follows:

$$CDF(s) = \frac{NumQMS{\leq}s}{TNQ}. \tag{7}$$

In this equation: *NumQMS* – number of queries with matching scores, *TNQ* – total *NumQMS*, s – the matching score threshold.

– **Combined Analysis:** By jointly assessing the CMC and CDF metrics, our evaluation provides a full overview of the proposed intelligent video surveillance system. The CMC curve offers insights into the system's ranking performance, while the CDF curve provides information on the distribution of matching scores. Together, these measures contribute to a thorough assessment of the system's capacity to accurately re-identify humans in real-world scenarios, incorporating both top-ranked matches and overall score distributions.

Our methodology directly addresses the difficulties indicated in the Introduction by leveraging the Market-1501 dataset to examine real-world performance. We utilize deep learning architectures, including VGG16 with triplet loss, and evaluate several distance measures to increase re-identification accuracy. The combination of CMC and CDF ensures a thorough performance evaluation, bridging existing information gaps and confirming our technique in the Results section.

3. Results and Discussion. In this chapter, we report the outcomes of our research on people re-identification approaches applying deep learning algorithms. We investigate these data comprehensively, providing insights into the strengths and limitations of each technique. The results are presented utilizing figures, graphs, and tables for clarity and simplicity of understanding.

3.1. Key Findings and Interpretation. Table 2 summarizes the Rank-1, Rank-5, and Rank-10 accuracy rates for each technique. Figures 3 to 12 provide visual representations of the results through CMC and CDF curves.

Table 2. Rank-1, Rank-5, and Rank-10 accuracy rates for each technique

Technique	Rank-1	Rank-5	Rank-10
Euclidean	2%	5.5%	12.5%
Cosine	0.30%	4.73%	10.21%
Mahalanobis	2%	5,88%	11.56%
Siamese	0%	8%	12.5%
Triplet	0.06%	78.6%	93%

Figure 3(a) illustrates the CMC curve for Euclidean Distance, which shows a gradual increase in identification probability across different ranks but exhibits poor Rank-1 accuracy (2%). While the identification rate improves at Rank-5 (5.5%) and Rank-10 (12.5%), it remains ineffective for real-world person re-identification applications. Figure 3(b) illustrates the Cumulative Distribution Function (CDF) curve of Euclidean distances, highlighting the distribution of matching scores obtained from our experiments. Similarly, Figures 4(a) and Figure 5(a) depict the CMC curves for Cosine and Mahalanobis distances, respectively. The Cosine distance metric achieves the lowest Rank-1 accuracy (0.30%) and struggles to distinguish individuals effectively at lower ranks, showing limited Rank-5 (4.73%) and Rank-10 (10.21%) accuracy. The Mahalanobis distance metric slightly outperforms both Euclidean and Cosine, achieving 2% Rank-1 accuracy, but its performance remains moderate with 5.88% at Rank-5 and 11.56% at Rank-10. The corresponding CDF curves in Figures 4(b) and Figure 5(b) further confirm these trends, demonstrating slow convergence toward higher identification probabilities.

Rank-1 accuracy measures the percentage of queries where the correct match appears as the first result, reflecting immediate recognition accuracy. Rank-5 accuracy measures the percentage where the correct match is within the top five results, indicating performance with a small candidate set. These differ from traditional accuracy metrics, like precision, which measure correct predictions out of all predictions without considering ranking.

In contrast, deep learning-based techniques show a notable improvement over traditional methods. Figure 6(a) presents the CMC curve for the Siamese network, which performs better than traditional distance metrics at higher ranks but fails at Rank-1 (0%). However, the method shows a significant improvement at Rank-5 (8%) and Rank-10 (12.5%), indicating better feature representation than distance-based metrics.

Figure 7(a) illustrates the CMC curve for the Triplet network, which outperforms all other techniques, achieving a Rank-1 accuracy of 0.06%. Despite this low Rank-1 accuracy, the Triplet network shows a dramatic increase at Rank-5 (78.6%) and Rank-10 (93%), demonstrating its superior ability to rank correct matches in higher ranks. The CDF curves in Figures 6 (b) and Figure 7 (b) reinforce these findings, highlighting the robustness of the Triplet network in learning discriminative embeddings for re-identification.

Figure 8 illustrates the average precision, recall, and F1-score for each technique, providing a broader evaluation beyond rank-based accuracy. The results highlight the Triplet network's superior performance across all ranking levels, confirming its ability to learn highly discriminative features and maximize correct match ranking at later ranks. In contrast, traditional distance metrics, as well as the Siamese network, exhibit significantly lower precision and recall, indicating their limited effectiveness in person re-identification tasks.

3.2. Comparison with Previous Studies. Our findings align with prior research emphasizing the superiority of deep metric learning approaches over traditional distance metrics. Compared to state-of-the-art results on Market-1501, studies such as [28] and [29] demonstrate that triplet loss significantly enhances person re-identification performance by optimizing the embedding space, particularly at higher rank levels (Rank-5 and Rank-10).

Our results confirm these observations, as the Triplet network achieves the highest accuracy at Rank-5 (78.6%) and Rank-10 (93%), demonstrating its effectiveness in ranking correct matches. However, the Triplet network shows weak Rank-1 accuracy (0.06%), suggesting that other deep learning approaches, such as Siamese networks, may still be useful at lower ranks.

Additionally, the study by [30] emphasizes that while deep learning methods significantly enhance identification accuracy, their effectiveness varies depending on model design and training strategy. In our case, the Triplet network demonstrates competitive performance at higher ranks

(Rank-5 and Rank-10), confirming its ability to learn meaningful embeddings for person re-identification. Although we utilized the VGG16 architecture – which is not the most recent development in CNN design – this choice was intentional to establish a clear and reproducible baseline. In future work, we plan to extend this framework by integrating more advanced backbones such as ResNet and EfficientNet, and subsequently applying transformer-based architectures to further enhance recognition performance.

Compared to traditional distance-based methods such as Euclidean, Cosine, and Mahalanobis metrics discussed in References [24, 25, 26], our approach demonstrates a significant improvement in Rank-5 and Rank-10 accuracy. However, compared to recent state-of-the-art deep learning methods listed in Table 1, our model offers competitive higher-rank performance but still lags in Rank-1 accuracy and overall end-to-end accuracy due to its reliance on an older backbone (VGG16).

3.3. Training Protocols. A summary of each technique's strengths and limitations is provided in Table 3.

Table 3. Summary of strengths and limitations for each technique based on quantitative metrics

Technique	Strengths	Limitations
Euclidean	Simple to implement and computationally efficient	Low accuracy at Rank-1 (2%) and Rank-5 (5.5%), slightly better than Cosine
Cosine	Effective for high-dimensional data	Very low Rank-1 accuracy (0.3%), lower Rank-5 accuracy (4.73%) than Euclidean
Mahalanobis	Accounts for correlations in data	Low Rank-1 (2%) and Rank-5 (5.88%), similar to Euclidean but computationally heavier
Siamese	Outperforms Euclidean and Cosine at Rank-5 (8%)	No correct matches at Rank-1 (0%), modest improvement at Rank-10 (12.5%)
Triplet	Significantly better accuracy at Rank-5 (78.6%) and Rank-10 (93%)	Requires complex training, very low Rank-1 accuracy (0.06%)

One unexpected result in our analysis was that the Cosine distance metric exhibited a moderate improvement at Rank-10 compared to Rank-1 and Rank-5. Specifically, its accuracy increased from 0.30% (Rank-1) to 4.73% (Rank-5) and 10.21% (Rank-10). This suggests that while Cosine distance struggles with immediate identification, it retains some useful feature representations that contribute to later-stage ranking improvements. However, since its overall accuracy remains lower than other methods, further investigation is needed to determine whether Cosine similarity can be effectively integrated with deep learning models to enhance ranking performance in re-identification tasks.

3.4. Impact of VGG16 Integration. Integrating VGG16 into the re-identification model significantly enhanced feature extraction by leveraging its deep hierarchical architecture. The convolutional layers of VGG16 captured fine-grained spatial details, improving the model's ability to distinguish individuals across different viewpoints, occlusions, and lighting conditions. As a result, deep learning-based techniques, particularly the Triplet network, achieved superior performance at Rank-5 (78.6%) and Rank-10 (93%), demonstrating their advantage over traditional distance metrics. However, the Triplet network exhibited low Rank-1 accuracy (0.06%), suggesting further improvements in early-stage recognition. Additionally, the increased computational complexity of VGG16 remains a challenge for real-time applications, highlighting the need for optimization techniques such as lighter architectures, model quantization, and feature distillation to balance accuracy and efficiency.

The integration of VGG16 for feature extraction follows the widely recognized approach by Simonyan and Zisserman (2014) [12], with our novel contribution lying in its application to the specific task of person re-identification, where we demonstrate its significant enhancement of identification accuracy compared to traditional methods.

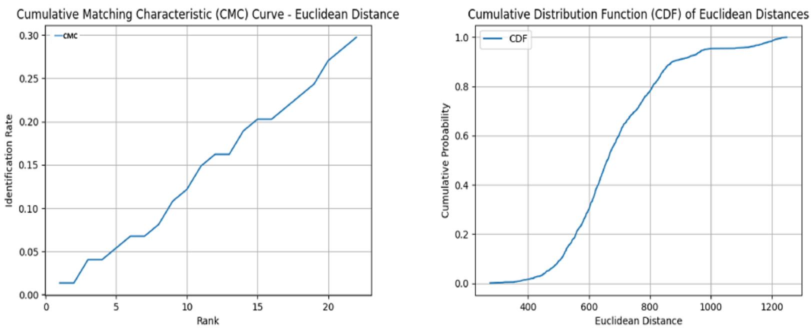

Fig. 3. a) (CMC) Curve – Euclidean Distance; b) (CDF) Curve – Euclidean Distance

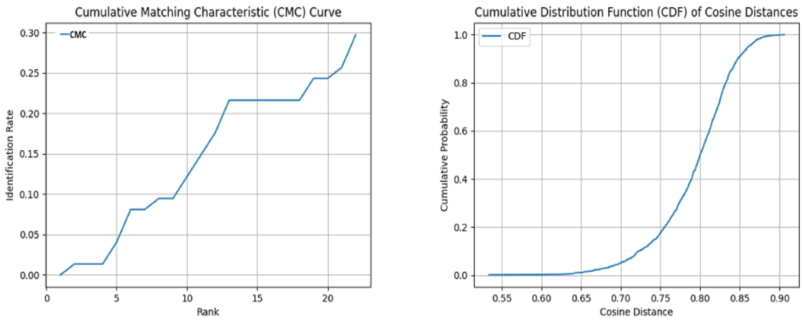

Fig. 4. a) (CMC) Curve Cosine Distance; b) (CDF) Curve Cosine Distance

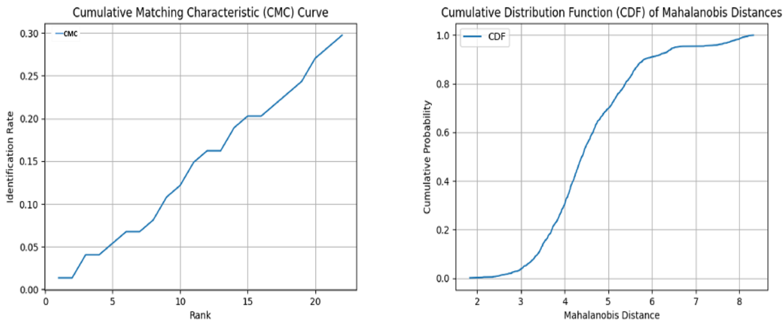

Fig. 5. a) (CMC) Curve Mahalanobis Distance; b) (CDF) Curve Mahalanobis Distance

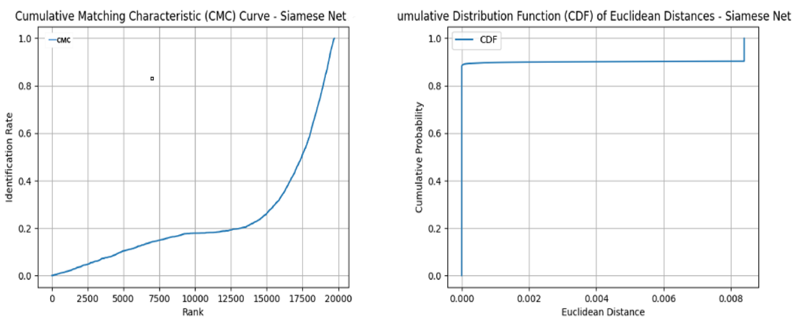

Fig. 6. a) CMC curve for Siamese networks; b) CDF curve for Siamese networks

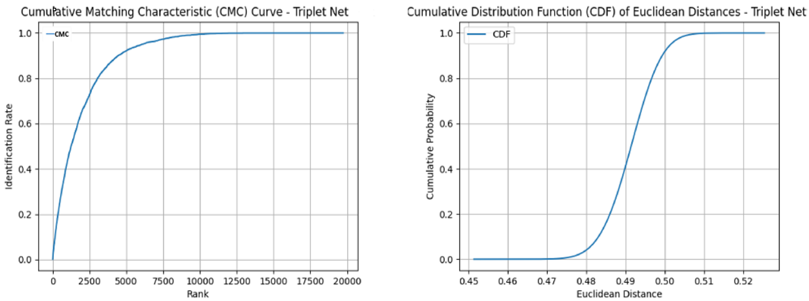

Fig. 7. a) CMC curve for Triplet networks; b) CDF curve for Triplet networks

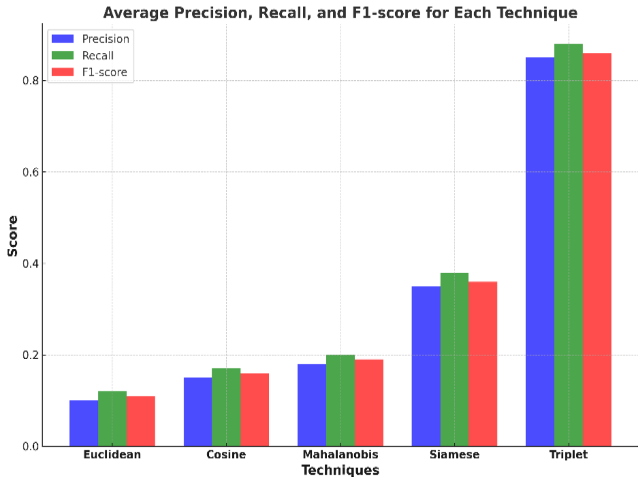

Fig. 8. Bar graph illustrating the average precision, recall, and F1-score for each technique

4. Conclusion and Future Directions. This paper provides a comparative analysis of traditional distance metrics and deep learning models for person re-identification, highlighting the advantages of deep metric learning approaches. Our findings demonstrate that while standard distance metrics such as Euclidean and Mahalanobis are computationally efficient, they suffer from significantly lower accuracy compared to deep learning-based techniques. The Triplet network, in particular, shows substantial gains at Rank-5 (78.6%) and Rank-10 (93%), making it highly effective for ranking-based retrieval tasks. However, its Rank-1 accuracy (0.06%) remains low, indicating limitations in immediate identification.

These results emphasize the importance of feature learning and embedding optimization in person re-identification. Deep learning models capture complex patterns and variations in appearance, making them more robust than traditional methods. However, the increasing computational complexity of deep networks, particularly Triplet loss-based models, poses challenges for real-time deployment. This highlights the need for optimization techniques, such as lightweight architectures, quantization, and knowledge distillation, to balance accuracy and efficiency.

Beyond re-identification accuracy, our findings contribute to broader discussions on the scalability and applicability of deep learning for intelligent video surveillance. Future research should focus on:

- Constructing hybrid models that balance efficiency and accuracy.
- Integrating domain adaptation techniques to improve cross-dataset generalization.
- Studying privacy-preserving mechanisms (e.g., federated learning, differential privacy) for ethical surveillance applications.

Ultimately, this study reinforces the growing role of deep learning in person re-identification, providing valuable insights for researchers and practitioners to develop scalable, efficient, and ethical surveillance systems. The advancements in this field have far-reaching implications, from public security and law enforcement to smart city applications, where accurate and efficient identification is critical for real-world deployment.

References

1. Ezzahoutz A., Youssef H.M., Thami R.O.H. Detection evaluation and testing region incoming people’s in a simple camera view. Second International Conference on the Innovative Computing Technology (INTECH 2012). Casablanca: IEEE, 2012. pp. 179–183. DOI: 10.1109/INTECH.2012.6457804.
2. Ezzahout A., Hadi Y., Thami R.O.H. Performance Evaluation of Mobile Person Detection and Area Entry Tests through a One-View Camera. Journal of Information. 2012. vol. 2. no. 3.
3. Sun Z., et al. A comprehensive review of pedestrian re-identification based on deep learning. Complex Intell. Syst. 2024. vol. 10. no. 2. pp. 1733–1768. DOI: 10.1007/s40747-023-01229-7.
4. Zahra A., Perwaiz N., Shahzad M., Fraz M.M. Person re-identification: A retrospective on domain specific open challenges and future trends. Pattern Recognition. 2023. vol. 142. p. 109669. DOI: 10.1016/j.patcog.2023.109669.
5. De Maesschalck R., Jouan-Rimbaud D., Massart D.L. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems. 2000. vol. 50. no. 1. pp. 1–18. DOI: 10.1016/S0169-7439(99)00047-7.
6. Ghorbani H. Mahalanobis distance and its application for detecting multivariate outliers. Facta Universitatis, Series: Mathematics and Informatics. 2019. pp. 583–595. DOI: 10.22190/FUMI1903583G.
7. Mclachlan G. Mahalanobis Distance. Resonance. 1999. vol. 4. pp. 20–26. DOI: 10.1007/BF02834632.

8. Tena S., Hartanto R., Ardiyanto I. Content-based image retrieval for fabric images: A survey. Indones. J. Electr. Eng. Comput. Sci. 2021. vol. 23. no. 3. DOI: 10.11591/ijeecs.v23.i3.pp1861-1872.
9. Samit Hatem A., Altememe M.S., Fadhel M.A. Identifying corn leaves diseases by extensive use of transfer learning: a comparative study. Indones. J. Electr. Eng. Comput. Sci. 2023. vol. 29. no. 2. DOI: 10.11591/ijeecs.v29.i2.pp1030-1038.
10. Gottipati S.B., Thumbur G. Multi-modal fusion deep transfer learning for accurate brain tumor classification using magnetic resonance imaging images. Indones. J. Electr. Eng. Comput. Sci. 2024. vol. 34. no. 2. DOI: 10.11591/ijeecs.v34.i2.pp825-834.
11. Sehree N.A., Khidhir A.M. Olive trees cases classification based on deep convolutional neural network from unmanned aerial vehicle imagery. Indones. J. Electr. Eng. Comput. Sci. 2022. vol. 27. no. 1. pp. 92–101. DOI: 10.11591/ijeecs.v27.i1.pp92-101.
12. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556. 2015. DOI: 10.48550/arXiv.1409.1556.
13. Aufar Y., Sitanggang I.S. Face recognition based on Siamese convolutional neural network using Kivy framework. IJEECS. 2022. vol. 26. no. 2. pp. 764–772. DOI: 10.11591/ijeecs.v26.i2.
14. Xu J. A deep learning approach to building an intelligent video surveillance system. Multimedia Tools and Applications. 2021. vol. 80(4). pp. 5495–5515. DOI: 10.1007/s11042-020-09964-6.
15. Alami M.I., Ez-Zahout A., Omary F. Enhanced people re-identification in cctv surveillance using deep learning: a framework for real-world applications. Informatics and Automation. 2025. vol. 24. no. 2. pp. 583–603. DOI: 10.15622/ia.24.2.8.
16. Moghaddam A.S., Anvari F., Haghighi M.M., Fakhari M., Mohammadi M.R. IUST_PersonReId: A New Domain in Person Re-Identification Datasets. arXiv preprint arXiv:2412.18874. 2024. DOI: 10.48550/arXiv.2412.18874.
17. Wang G.A, Gong S., Cheng J., Hou Z. Faster Person Re-identification. European conference on computer vision (ECCV 2020). Cham: Springer International Publishing, 2020. vol. 12353. pp. 275–292. DOI: 10.1007/978-3-030-58598-3_17.
18. Rami H., Ospici M., Lathuilière S. Online Unsupervised Domain Adaptation for Person Re-identification. arXiv preprint arXiv:2205.04383. 2022. DOI: 10.48550/arXiv.2205.04383.
19. Khan S.U., Khan N., Hussain T., Baik S.W. An intelligent correlation learning system for person Re-identification. Engineering Applications of Artificial Intelligence. 2024. vol. 128. DOI: 10.1016/j.engappai.2023.107213.
20. Khan S.U., Hussain T., Ullah A., Baik S.W. Deep-ReID: deep features and autoencoder assisted image patching strategy for person re-identification in smart cities surveillance. Multimed Tools Appl. 2024. vol. 83. no. 5. pp. 15079–15100. DOI: 10.1007/s11042-020-10145-8.
21. He S., Luo H., Wang P., Wang F., Li H., Jiang W. TransReID: Transformer-Based Object Re-Identification (CVPR). Proceedings of the IEEE/CVF international conference on computer vision. 2023. pp. 15013–15022.
22. Ye M., Chen S., Li C., Zheng W.-S., Crandall D., Du B. Transformer for Object Re-Identification: A Survey. arXiv preprint arXiv:2401.06960. 2024. DOI: 10.48550/arXiv.2401.06960.
23. Choudhary A., Mishra D., Karmakar A. Domain Adaptive Egocentric Person Re-identification. Computer Vision and Image Processing: 5th International Conference (CVIP 2020). 2021. vol. 1378. pp. 81–92. DOI: 10.1007/978-981-16-1103-2_8.

24. Taye M.M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. Computers. 2023. vol. 12. no. 5. DOI: 10.3390/computers12050091.
25. Chopra S., Hadsell R., LeCun Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA: IEEE, 2005. vol. 1. pp. 539–546. DOI: 10.1109/CVPR.2005.202.
26. Schroff F., Kalenichenko D., Philbin J. FaceNet: A unified embedding for face recognition and clustering. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. pp. 815–823. DOI: 10.1109/CVPR.2015.7298682.
27. Karanam S., Gou M., Wu Z., Rates-Borras A., Camps O., Radke R.J. A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets. arXiv preprint arXiv:1605.09653. 2018.
28. Xu P., Zhu X. DeepChange: A Long-Term Person Re-Identification Benchmark with Clothes Change. IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023. pp. 11162–11171. DOI: 10.1109/ICCV51070.2023.01028.
29. Ye M., Shen J., Lin G., Xiang T., Shao L., Hoi S.C. Deep Learning for Person Re-identification: A Survey and Outlook. arXiv preprint arXiv:2001.04193. 2021. DOI: 10.48550/arXiv.2001.04193.
30. Ning E., Wang C., Zhang H., Ning X., Tiwari P. Occluded person re-identification with deep learning: A survey and perspectives. Expert Systems with Applications. 2024. vol. 239. DOI: 10.1016/j.eswa.2023.122419.

Idrissi Alami Mossaab — Ph.d. student, Department of computer science/faculty of sciences, Mohammed V University in Rabat. Research interests: big data engineering, information systems administration, computer engineering. The number of publications — 1. mossaab_idrissialami@um5.ac.ma; Av. des Nations Unies, 10000, Rabat, Morocco; office phone: +212(661)065-800.

Ez-zahout Abderrahmane — Professor of the department, Department of computer science/faculty of sciences, Mohammed V University in Rabat. Research interests: computer sciences, digital systems, big data, computer vision, intelligent systems. The number of publications — 30. a.ezzahout@um5r.ac.ma; Av. des Nations Unies, 10000, Rabat, Morocco; office phone: +212(669)455-252.

Omary Fouzia — Leader ipss team research, professor of the department, Department of computer science/faculty of sciences, Mohammed V University in Rabat. Research interests: cybersecurity, blockchain, cryptocurrency. The number of publications — 59. omary@fsr.ac.ma; Av. des Nations Unies, 10000, Rabat, Morocco; office phone: +212(661)391-420.

УДК 006.72 **DOI** 10.15622/ia.24.3.9

М. Идрисси Алами, А. Эз-захут, Ф. Омари

СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДОВ ПОВТОРНОЙ ИДЕНТИФИКАЦИИ ЛИЧНОСТИ НА ОСНОВЕ МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ

Идрисси Алами М., Эз-захут А., Омари Ф. **Сравнительное исследование методов повторной идентификации личности на основе моделей глубокого обучения.**

Аннотация. Повторная идентификация личности (Re-ID) имеет ключевую роль в системах интеллектуального видеонаблюдения, требуя точного распознавания людей с нескольких точек обзора камеры. Традиционные методы, основанные на метриках расстояния (евклидово и косинусное), сталкиваются с трудностями при обработке вариаций поз и случаев окклюзии, что ограничивает их эффективность. В данном исследовании рассматриваются модели глубокого метрического обучения, в частности сиамские и триплетные сети, для повышения точности повторной идентификации. Мы оцениваем эти методы на наборе данных Market-1501 с использованием кривых кумулятивной характеристики соответствия (CMC) и кумулятивной функции распределения (CDF). Результаты показывают, что триплетная сеть превосходит традиционные подходы на более высоких рангах, достигая точности 78,6% на Rank-5 и точности 93% на Rank-10, при этом демонстрируя низкую точность на Rank-1 (0,06%). Для сравнения, методы на основе евклидова и косинусного расстояний показывают низкую производительность на Rank-1 (2% и 0,30% соответственно), что подчеркивает их ограничения. Кроме того, включение архитектуры VGG16 улучшает извлечение признаков, повышая эффективность распознавания за счет улавливания мельчайших пространственных деталей. Данное сравнительное исследование показывает эффективность методов глубокого метрического обучения и подчеркивает его потенциал для практического применения в системах видеонаблюдения. Однако вычислительные требования глубоких сетей создают сложности для работы в реальном времени. Будущие исследования должны быть сосредоточены на оптимизации эффективности модели, снижении вычислительных затрат и тестировании в реальном времени.

Ключевые слова: метрики CMC/CDF, сверточные нейронные сети (CNN), глубокое обучение, повторная идентификация личности, VGG16, системы видеонаблюдения.

Литература

1. Ezzahoutz A., Youssef H.M., Thami R.O.H. Detection evaluation and testing region incoming people's in a simple camera view. Second International Conference on the Innovative Computing Technology (INTECH 2012). Casablanca: IEEE, 2012. pp. 179–183. DOI: 10.1109/INTECH.2012.6457804.
2. Ezzahout A., Hadi Y., Thami R.O.H. Performance Evaluation of Mobile Person Detection and Area Entry Tests through a One-View Camera. Journal of Information. 2012. vol. 2. no. 3.
3. Sun Z., et al. A comprehensive review of pedestrian re-identification based on deep learning. Complex Intell. Syst. 2024. vol. 10. no. 2. pp. 1733–1768. DOI: 10.1007/s40747-023-01229-7.
4. Zahra A., Perwaiz N., Shahzad M., Fraz M.M. Person re-identification: A retrospective on domain specific open challenges and future trends. Pattern Recognition. 2023. vol. 142. p. 109669. DOI: 10.1016/j.patcog.2023.109669.

5. De Maesschalck R., Jouan-Rimbaud D., Massart D.L. The Mahalanobis distance. Chemometrics and Intelligent Laboratory Systems. 2000. vol. 50. no. 1. pp. 1–18. DOI: 10.1016/S0169-7439(99)00047-7.
6. Ghorbani H. Mahalanobis distance and its application for detecting multivariate outliers. Facta Universitatis, Series: Mathematics and Informatics. 2019. pp. 583–595. DOI: 10.22190/FUMI1903583G.
7. Mclachlan G. Mahalanobis Distance. Resonance. 1999. vol. 4. pp. 20–26. DOI: 10.1007/BF02834632.
8. Tena S., Hartanto R., Ardiyanto I. Content-based image retrieval for fabric images: A survey. Indones. J. Electr. Eng. Comput. Sci. 2021. vol. 23. no. 3. DOI: 10.11591/ijeecs.v23.i3.pp1861-1872.
9. Samit Hatem A., Altememe M.S., Fadhel M.A. Identifying corn leaves diseases by extensive use of transfer learning: a comparative study. Indones. J. Electr. Eng. Comput. Sci. 2023. vol. 29. no. 2. DOI: 10.11591/ijeecs.v29.i2.pp1030-1038.
10. Gottipati S.B., Thumbur G. Multi-modal fusion deep transfer learning for accurate brain tumor classification using magnetic resonance imaging images. Indones. J. Electr. Eng. Comput. Sci. 2024. vol. 34. no. 2. DOI: 10.11591/ijeecs.v34.i2.pp825-834.
11. Sehree N.A., Khidhir A.M. Olive trees cases classification based on deep convolutional neural network from unmanned aerial vehicle imagery. Indones. J. Electr. Eng. Comput. Sci. 2022. vol. 27. no. 1. pp. 92–101. DOI: 10.11591/ijeecs.v27.i1.pp92-101.
12. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv preprint arXiv:1409.1556. 2015. DOI: 10.48550/arXiv.1409.1556.
13. Aufar Y., Sitanggang I.S. Face recognition based on Siamese convolutional neural network using Kivy framework. IJEECS. 2022. vol. 26. no. 2. pp. 764–772. DOI: 10.11591/ijeecs.v26.i2.
14. Xu J. A deep learning approach to building an intelligent video surveillance system. Multimedia Tools and Applications. 2021. vol. 80(4). pp. 5495–5515. DOI: 10.1007/s11042-020-09964-6.
15. Alami M.I., Ez-Zahout A., Omary F. Enhanced people re-identification in cctv surveillance using deep learning: a framework for real-world applications. Informatics and Automation. 2025. vol. 24. no. 2. pp. 583–603. DOI: 10.15622/ia.24.2.8.
16. Moghaddam A.S., Anvari F., Haghighi M.M., Fakhari M., Mohammadi M.R. IUST_PersonReId: A New Domain in Person Re-Identification Datasets. arXiv preprint arXiv:2412.18874. 2024. DOI: 10.48550/arXiv.2412.18874.
17. Wang G.A, Gong S., Cheng J., Hou Z. Faster Person Re-identification. European conference on computer vision (ECCV 2020). Cham: Springer International Publishing, 2020. vol. 12353. pp. 275–292. DOI: 10.1007/978-3-030-58598-3_17.
18. Rami H., Ospici M., Lathuilière S. Online Unsupervised Domain Adaptation for Person Re-identification. arXiv preprint arXiv:2205.04383. 2022. DOI: 10.48550/arXiv.2205.04383.
19. Khan S.U., Khan N., Hussain T., Baik S.W. An intelligent correlation learning system for person Re-identification. Engineering Applications of Artificial Intelligence. 2024. vol. 128. DOI: 10.1016/j.engappai.2023.107213.
20. Khan S.U., Hussain T., Ullah A., Baik S.W. Deep-ReID: deep features and autoencoder assisted image patching strategy for person re-identification in smart cities surveillance. Multimed Tools Appl. 2024. vol. 83. no. 5. pp. 15079–15100. DOI: 10.1007/s11042-020-10145-8.

21. He S., Luo H., Wang P., Wang F., Li H., Jiang W. TransReID: Transformer-Based Object Re-Identification (CVPR). Proceedings of the IEEE/CVF international conference on computer vision. 2023. pp. 15013–15022.
22. Ye M., Chen S., Li C., Zheng W.-S., Crandall D., Du B. Transformer for Object Re-Identification: A Survey. arXiv preprint arXiv:2401.06960. 2024. DOI: 10.48550/arXiv.2401.06960.
23. Choudhary A., Mishra D., Karmakar A. Domain Adaptive Egocentric Person Re-identification. Computer Vision and Image Processing: 5th International Conference (CVIP 2020). 2021. vol. 1378. pp. 81–92. DOI: 10.1007/978-981-16-1103-2_8.
24. Taye M.M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. Computers. 2023. vol. 12. no. 5. DOI: 10.3390/computers12050091.
25. Chopra S., Hadsell R., LeCun Y. Learning a Similarity Metric Discriminatively, with Application to Face Verification. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). San Diego, CA, USA: IEEE, 2005. vol. 1. pp. 539–546. DOI: 10.1109/CVPR.2005.202.
26. Schroff F., Kalenichenko D., Philbin J. FaceNet: A unified embedding for face recognition and clustering. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015. pp. 815–823. DOI: 10.1109/CVPR.2015.7298682.
27. Karanam S., Gou M., Wu Z., Rates-Borras A., Camps O., Radke R.J. A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets. arXiv preprint arXiv:1605.09653. 2018.
28. Xu P., Zhu X. DeepChange: A Long-Term Person Re-Identification Benchmark with Clothes Change. IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, 2023. pp. 11162–11171. DOI: 10.1109/ICCV51070.2023.01028.
29. Ye M., Shen J., Lin G., Xiang T., Shao L., Hoi S.C. Deep Learning for Person Re-identification: A Survey and Outlook. arXiv preprint arXiv:2001.04193. 2021. DOI: 10.48550/arXiv.2001.04193.
30. Ning E., Wang C., Zhang H., Ning X., Tiwari P. Occluded person re-identification with deep learning: A survey and perspectives. Expert Systems with Applications. 2024. vol. 239. DOI: 10.1016/j.eswa.2023.122419.

Идрисси Алами Моссааб — аспирант, кафедра компьютерных наук/факультет естественных наук, Университет Мохаммеда V в Рабате. Область научных интересов: инженерия больших данных, администрирование информационных систем, компьютерная инженерия. Число научных публикаций — 1. mossaab_idrissialami@um5.ac.ma; проспект Объединенных Наций, 10000, Рабат, Марокко; р.т.: +212(661)065-800.

Эз-захут Абдеррахман — профессор кафедры, кафедра компьютерных наук/факультет естественных наук, Университет Мохаммеда V в Рабате. Область научных интересов: компьютерные науки, цифровые системы, большие данные, компьютерное зрение, интеллектуальные системы. Число научных публикаций — 30. a.ezzahout@um5r.ac.ma; проспект Объединенных Наций, 10000, Рабат, Марокко; р.т.: +212(669)455-252.

Омари Фузия — руководитель исследовательской группы IPSS, профессор кафедры, кафедра компьютерных наук/факультет естественных наук, Университет Мохаммеда V в Рабате. Область научных интересов: кибербезопасность, блокчейн, криптовалюта. Число научных публикаций — 59. omary@fsr.ac.ma; проспект Объединенных Наций, 10000, Рабат, Марокко; р.т.: +212(661)391-420.