DOI: 10.24412/2181-144X-2026-1-41-53

Babadjanov E., Serjanova D., Faizullayeva M.

MODERN APPROACHES TO ONLINE MULTI-OBJECT SURVEYS BASED ON DEEP METRIC LEARNING AND RE-IDENTIFICATION

Elmurod Babadjanov[1][0000-0002-5554-6727], ***Dilbar Serjanova***[2][0009-0006-7138-8112], ***Manzura Faizullayeva***[3][0009-0000-9816-5548],

[1]*Head of the Department of Cybersecurity Engineering at Nukus State Technical University, DSc, E-mail: elmurbes@gmail.com*
[2]*Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, doctoral student of the department of SAP, E-mail: dilyaserjanova@gmail.com*
[3]*Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, doctoral student of the department of SAP, E-mail:fayzullaeva7117@gmail.com*

Abstract. This article analyzes modern approaches in the field of Multiple Object Tracking (MOT) based on deep metric learning and re-identification (Re-ID). The main focus is on solving the problem of heavy occlusion, unsupervised Re-ID, Siamese neural networks and metric learning methods. The analysis was conducted based on five high-impact scientific works, including approaches such as the unsupervised Re-ID module, the Siamese Deep Metric Tracker (SDMT), metric learning based on the von Mises-Fisher (vMF) distribution, and SiamMOT. The results show that appearance-based embeddings (feature vectors based on the visual appearance of the object) and occlusion estimation modules help correctly store object identifiers even under conditions of heavy occlusion, significantly reducing identifier switching (ID switch). These approaches make it possible to accurately assess the similarity of objects and re-identify objects lost under a temporary barrier. However, maintaining a balance between high accuracy and real-time performance, as well as saving computing resources, remains a pressing issue.

Keywords: Multi-Object Tracking, Re-Identification, Deep Metric Learning, Siamese Networks, Occlusion Handling, Unsupervised Learning, Online Tracking.

Annotatsiya. Ushbu maqolada chuqur metrik oʻrganish va qayta identifikatsiyalash (Re-ID) asosida koʻp obyektli kuzatuv (MOT) sohasidagi zamonaviy yondashuvlar tahlil qilingan. Asosiy e’tibor ogʻir okklyuziya, nazoratsiz Re-ID, Siam neyron tarmoqlari va metrik oʻrganish usullari muammosini hal qilishga qaratilgan. Tahlil beshta yuqori ta’sirga ega ilmiy ishlar, jumladan, nazorat qilinmaydigan Re-ID moduli, Siamese Deep Metric Tracker (SDMT), von Mises-Fisher (vMF) taqsimotiga asoslangan metrik oʻrganish va SiamMOT kabi yondashuvlar asosida oʻtkazildi. Natijalar shuni koʻrsatadiki, tashqi koʻrinishga asoslangan joylashtirishlar (obyektning vizual koʻrinishiga asoslangan xususiyat vektorlari) va okklyuziyani baholash modullari obyekt identifikatorlarini hatto kuchli okklyuziya sharoitida ham toʻgʻri saqlashga yordam beradi, bu esa identifikatorlarni almashtirishni (ID switch) sezilarli darajada kamaytiradi.

Ushbu yondashuvlar obyektlarning oʻxshashligini aniq baholash va vaqtinchalik toʻsiq ostida yoʻqolgan obyektlarni qayta aniqlash imkonini beradi. Biroq, yuqori aniqlik va real vaqt rejimidagi unumdorlik oʻrtasidagi muvozanatni saqlash hamda hisoblash resurslarini tejash dolzarb muammo boʻlib qolmoqda.

Kalit soʻzlar: koʻp obyektli kuzatuv, qayta aniqlash, chuqur metrik oʻrganish, Siam tarmoqlari, okklyuziya bilan ishlash, nazoratsiz oʻrganish, onlayn kuzatuv.

Аннотация. В данной статье анализируются современные подходы в области мониторинга множественных объектов (MOT), основанные на глубоком метрическом обучении и повторной идентификации (Re-ID). Основное внимание уделяется решению проблем тяжелой окклюзии, неконтролируемого Re-ID, сиамских нейронных сетей и метрических методов обучения. Анализ был проведен на основе пяти высокоэффективных научных работ, включая такие подходы, как неконтролируемый модуль Re-ID, Siamese Deep Metric Tracker (SDMT), метрическое обучение на основе распределения Мизеса-Фишера (vMF) и SiamMOT. Результаты показывают, что встраивания на основе внешнего вида (векторы признаков, основанные на визуальном виде объекта) и модули оценки окклюзии помогают правильно хранить идентификаторы объектов даже в условиях сильной окклюзии, значительно снижая переключение идентификаторов (ID switch).

Эти подходы позволяют точно оценить сходство объектов и повторно идентифицировать объекты, потерянные под временным барьером. Однако сохранение баланса между высокой точностью и производительностью в режиме реального времени, а также экономия вычислительных ресурсов остается актуальной проблемой.

Ключевые слова: Отслеживание нескольких объектов, повторная идентификация, глубокое метрическое обучение, сиамские сети, обработка окклюзий, обучение без контроля, онлайн-отслеживание.

Introduction

200 scientific articles were analyzed to form the theoretical basis of this research. The study focused on identifying advanced approaches to the Re-ID process and related topics in multi-object tracking systems. The articles were analyzed for content consistency, effectiveness of the proposed methods, mechanisms for restoring occlusion, and opportunities for unsupervised learning. As a result, 5 main articles were selected aimed at improving the Re-ID process in multi-object tracking systems and ensuring stable tracking in occlusion conditions.

Multiple Object Tracking consists of identifying and tracking the trajectories of several objects simultaneously in video streams, which is one of the most complex tasks of computer vision. In recent years, Deep Learning technologies, particularly deep metric learning (deep metric learning) and Re-ID methods, have significantly increased the efficiency of MOT.

This article analyzes five important research papers focused on the main challenges of the MOT problem - heavy occlusion, ID switch, and real-time operation. The general idea of these works is to improve the quality of observation by creating robust embeddings (feature vectors) based on the visual appearance of objects and linking them over time. These approaches make it possible to accurately assess the similarity of objects, re-identify objects lost under a temporary barrier, and ensure the stability of the observation process.

Main part

The literature selection process in this study was organized in accordance with the international PRISMA 2020 standards. At the initial identification stage, a total of 200 articles were identified from various scientific databases. In the subsequent stages, these articles were gradually sorted based on screening and eligibility criteria, and relevant scientific works were selected for final analysis. Before the screening, 60 duplicate articles were excluded and 140 articles were reviewed. At the screening stage, 90 articles were excluded. 50 articles were selected for full-text evaluation, 5 of which were rejected. The remaining 45 articles were evaluated based on criteria, and 40 were excluded for various reasons. In the end, 5 articles fully met the selection criteria and were included in the systematic review. No new articles found from additional sources. The PRISMA diagram (Fig. 1) shows the identification, screening, match assessment, and final input process.

The articles obtained as a result of indexing were subjected to in-depth expert analysis regarding content consistency, the effectiveness of the proposed methods, mechanisms for restoring occlusion that arises during the observation process, as well as the possibilities of unsupervised learning using the Re-ID model. In this process, special attention was paid to scientific work on creating similarity metrics based on deep metric learning, increasing tracker stability using Siamese architecture, and applying lightweight models for online tracking.

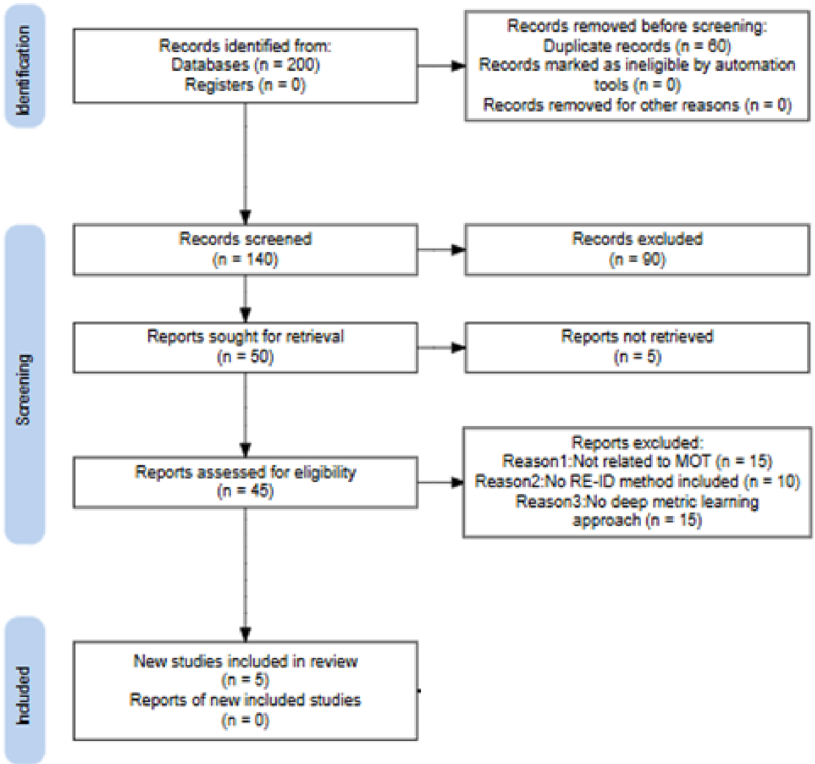

Fig.1 PRISMA 2020 flow diagram of study selection.

As a result of a multi-stage filter and scientific analysis, 5 main articles were selected from 200 articles that could provide the most methodologically appropriate and strong theoretical basis for future research. These articles were used as a scientific basis for improving the Re-ID process in multi-object tracking systems, maintaining stable tracking under occlusion conditions, and demonstrating the practical effectiveness of unsupervised learning models. The five main scientific sources used in this study are listed below.

In the study conducted at the University of Science and Technology of China, the issue of occlusion was identified as a common challenge in large-scale object tracking. This phenomenon often leads to incomplete detection of certain objects, resulting in decreased accuracy and stability of tracking. MOT is widely applied in re-identifying lost objects, a method that can potentially enhance tracking performance but requires labeled data for model training. Nevertheless, this approach proves ineffective in severe occlusion scenarios, where objects fail to be detected by the detector [1,7]. To address this, the study proposes two novel modules specifically designed for online MOT: an unsupervised Re-ID learning module and an occlusion estimation module. Contemporary online MOT systems predominantly rely on Re-ID modules based on detector-generated features; however, two major challenges remain:

-Traditional supervised approaches for Re-ID require ID labels for each individual or object. This leads to dimensional issues when handling a large number of identifiers, making adaptation to real-world large-scale data challenging.

- One of the most common problems is occlusion between objects, which can cause detector errors, object loss, and consequently, a decrease in tracking quality.

The study then provides a more detailed examination of the two proposed modules for online MOT.

1.Unsupervised Re-ID learning.

The unsupervised Re-ID learning module does not require labeled identification data (including pseudo labels) and thus avoids scalability issues. Learning Re-ID features without identifier tags implies that the system must be capable of extracting robust representations directly from video frames or image collections without explicit identity annotations. Objects appearing in adjacent frames typically share similar visual characteristics. Based on this observation, a matching-based loss function, analogous to contrastive or softmax approaches, is introduced. This matching-based loss (1) is employed to correctly associate objects across consecutive frames by assigning higher similarity scores to the most corresponding object in the next (or previous) frame, while penalizing mismatches with lower similarity scores.

$$L_{match} = -\sum_i log \frac{\exp\left(\frac{sim\left(f_i^t, f_{j^+}^{t+1}\right)}{\tau}\right)}{\sum_k \exp\left(\frac{sim\left(f_i^t, f_k^{t+1}\right)}{\tau}\right)} \quad (1)$$

where denotes the feature vector of object *i* in frame *t*, representing the "digital signature" of the object in that frame corresponds to the positive match in frame *t+1*, i.e., the representation of the same object in the subsequent frame; denotes other objects in frame *t+1* (negative samples). The function sim (·) measures the similarity between two vectors, typically implemented as cosine similarity. The parameter is a scalar temperature coefficient controlling the sharpness of the softmax distribution. The use of logarithm and softmax (the fraction above) follows a classical classification scheme, assigning the highest probability to the correct matching object while suppressing probabilities for incorrect matches.

Using formula (1), an object in the subsequent frame can be re-identified by assigning higher scores to correct pairs (positives) and lower scores to incorrect pairs (negatives). As a result, tracking accuracy is improved. This loss function is specifically designed to ensure that the system consistently recognizes the same object across consecutive frames.

2. Occlusion Estimation Module.

The occlusion estimation module predicts regions in which objects are likely to undergo occlusion and is utilized to assess the state of objects that remain undetected by the detector. Its primary function is to forecast areas within a frame where occlusions (overlaps) between objects may occur and, based on this information, to facilitate the recovery of objects missed by the detector [9]. The module is integrated as an additional parallel head into existing detectors (e.g., CenterNet/FairMOT), sharing common feature representations. It consists of two main components:

a) Occlusion Center Heatmap (with sigmoid activation). The sigmoid function maps the output values into the range [0,1], where each pixel represents the probability of being the center of an occlusion.

b) Occlusion Offset Map (without activation). In this case, the output values may be either negative or positive (e.g., -3.5 or 2.7). Therefore, no activation function is applied, allowing the model to predict unrestricted values as required.

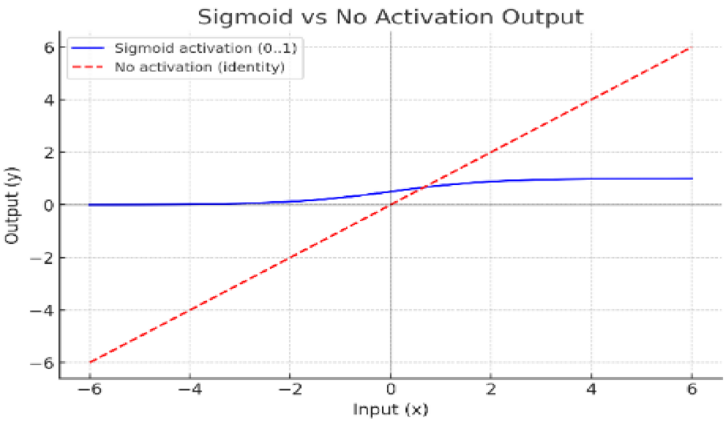

Fig.2. Sigmoid and identity activation function curves.

In the illustration below, the blue curve represents the sigmoid activation, which compresses any input into the [0,1] interval (interpreted as probability), while the red curve represents the identity (no activation), where the output retains the same sign and magnitude as the input. Consequently, the heatmap employs sigmoid activation to obtain probabilities, whereas the offset map omits activation since coordinate displacements must be preserved.

The research findings demonstrate that this approach outperforms several state-of-the-art MOT methods, and the proposed occlusion module significantly enhances the overall effectiveness of both Re-ID and tracking.

Michael Thoreau's work enhances MOT using deep metric learning. In this approach, deep metric learning (DML) is applied to improve the MOT problem in real-time scenarios. The method employs an appearance metric derived from a Siamese convolutional neural network (CNN) that was pre-trained on a large-scale "person Re-ID" dataset. A Siamese network is a model consisting of two identical neural networks that share the same weights. For example, given two input images, both pass through the network, and at the output, feature vectors are obtained for each image. In this case, the Siamese network generates 128-dimensional feature vectors (embeddings) [3,6].

1)Assignment - In each frame, the camera detects objects. The newly detected objects are then compared with the existing tracks. At this stage, both the object's motion (estimated using a Kalman filter) and its appearance (features extracted from the image) are considered. This ensures that each object is correctly linked to its corresponding track.

2)Detection Boosting - Sometimes the camera fails to detect an object (for instance, when a person is temporarily occluded by a shadow or another individual). In such cases, the system predicts the possible location of the object and searches in that region. If the predicted appearance matches, the object remains in the tracking process without being lost.

3)Tracklet Association - When an object temporarily disappears and then reappears, the system reconnects it to its previous track. Here, both motion consistency and visual appearance are taken into account. This reduces the risk of "ID switches" between objects.

SDMT support prevents objects in camera tracking from being "lost" or "mistakenly switching to another ID". The proposed SDMT method incorporates deep metric learning into MOT, achieving high accuracy at real-time speed and demonstrating good results on the MOT16 dataset. Appearance embeddings are useful for assignment and tracklet merging. However, boosting should be applied cautiously.

In recent years, one of the pressing research problems in the field of computer vision has been person re-identification in rapidly developing intelligent surveillance networks for public safety. Researchers at Sichuan University in China have conducted studies on several deep learning-based person re-identification methods.

The goal of person re-identification research is to track the same individuals across different cameras. However, traditional person re-identification methods require extensive manual labeling of tracked individuals, which is labor-intensive. With the widespread application of deep neural networks, many Re-ID methods based on deep learning for identifying multiple individuals have emerged[4].

Therefore, this article is aimed at facilitating the understanding of recent research results and future trends in the field. First, both classical and state-of-the-art deep learning-based person Re-ID methods are systematically classified. Second, a multidimensional taxonomy has been proposed that categorizes current deep learning-based person Re-ID methods into four main classes according to metric learning and representation learning.

A. Deep Metric Learning. This model studies the distance between images in such a way that images of the same person are located closer to each other, while images of different people are located further away. The basic triplet loss formula (3) is expressed as follows:

$$L_{triplet} = max(0, d(a,p) - d(a,n) + margin) \quad (2)$$

where- anchor (the main image), p – positive (another image of the same person), n – negative (an image of a different person), and d(·,·) – the distance function (e.g., Euclidean distance or cosine similarity). If the condition d(a,p) < d(a,n) holds in the model, then during airport surveillance, when the same individual is observed under different cameras, triplet loss ensures that all their images are grouped together in a close cluster.

If the model holds, then in airport surveillance, when the same person is captured under different cameras, triplet loss ensures that all their images are aggregated into a close cluster.

B. Local Feature Learning. Global embedding is not always sufficient, as a person's clothing, body pose, or partial occlusion can negatively affect accuracy. Therefore, the image is divided into parts (e.g., head, torso, legs), and features from each part are learned separately. For example, in a shopping mall, if a person is carrying a bag, the global embedding may become misleading. However, local features (such as clothing colors, face, or leg position) can preserve accurate identification.

C. GAN-based Learning. In Re-ID, GANs (Generative Adversarial Networks) help mitigate the problem of limited data. Using GANs, synthetic images of a person can be generated under different conditions (new viewpoints, varying lighting, or alternative clothing). One of the key advantages is that the model adapts to diverse appearances, reducing overfitting. For instance, in city surveillance where camera data is insufficient, GANs generate person images from new angles, which helps the model operate more effectively.

D. Sequence Feature Learning. Re-ID is often performed using only a single image. However, video streams contain a large amount of temporal information. Utilizing a sequence of multiple frames improves identification accuracy. One of the advantages is that it captures variations over time (such as walking style, direction, and speed). For example, in a stadium where a person's face may be difficult to recognize, their walking pattern and body movements can be used for identification.

By combining these four approaches, it is possible to achieve more accurate and reliable identification of individuals under various conditions. In addition, the advantages and disadvantages of these four categories, distinguished by their methodologies and motivations, are highlighted. As a result, certain challenges of person Re-ID and potential research directions are discussed [2].

In research conducted by researchers from Amazon Web Services, led by Bing Shuai's team, the focus was placed on improving online multi-object tracking (MOT). In conventional approaches, objects are first detected by a detector, and then association (matching) algorithms track each object's identity. However, such two-stage methods require substantial computational resources and face challenges in real-time performance. The authors propose a novel method called SiamMOT, which is based on a Siamese network architecture and jointly performs object detection and re-identification. Consequently, detection and tracking are executed within a single neural network framework [6].

The Siamese architecture is a dual-input neural network that compares appearances of the same object across different frames to determine whether they correspond to the same identity. For instance, if a pedestrian appears in one frame and reappears in another frame at a different location, the network compares both instances and confirms that they belong to the same individual.

The object detector localizes objects in each frame of the video through bounding boxes. The object detection loss function is defined as follows:

$$L_{det} = L_{cls} + L_{reg} \quad (3)$$

where L_{cls}- - the error in object class prediction, L_{reg}- the error in determining the object's coordinates. This formula ensures the proper functioning of the detector.

Re-ID features represent the visual characteristics of objects (e.g., clothing color, shape). The Re-ID loss function is:

$$L_{reid} = -log\frac{e^{s(x_i,x_j)}}{\sum_k e^{s(x_i x_k)}}, \tag{4}$$

where - the similarity between the features of two objects.

Using this formula, the network brings features of the same object closer together while pushing features of different objects farther apart. The overall loss function:

$$L = L_{det} + \lambda L_{reid} \tag{5}$$

In this approach, the detection and re-identification tasks are jointly optimized. The coefficient λ ensures a balance between the two tasks, helping to prevent object identity switches over time.

The SiamMOT model includes a motion modeling component, which evaluates object movement between consecutive frames to maintain continuous identification. To analyze the impact of motion modeling on tracking performance, two variants of Siam are proposed: one models motion implicitly, while the other explicitly models motion.

SiamMOT demonstrates high efficiency, capable of processing 720P videos at an average speed of 17 frames per second on a modern single GPU. This performance is sufficient for near real-time operation, enabling practical deployment in surveillance systems. The SiamMOT method is effective for object tracking because detection and re-identification are performed simultaneously. This reduces computational costs and allows deployment in real-time systems. The approach is not only useful for video surveillance but also for traffic monitoring, security cameras, and even livestock tracking in agriculture.

The general tracking association stage is one of the final and most important stages of the MOT system. This stage is shown in detail in the diagram below (Fig. 3). At the tracking association stage, the motion prediction calculated from previous frames, the object's reliability score, and the visual features are analyzed together. As a result, objects identified in the current frame are correctly linked to previous trajectories, new trajectories are created, and identifier exchange is prevented. This stage achieves particularly high efficiency when used in combination with methods such as Siamese networks, Kalman filter, and Hungarian algorithm, ensuring long-term stability of multi-object observation.

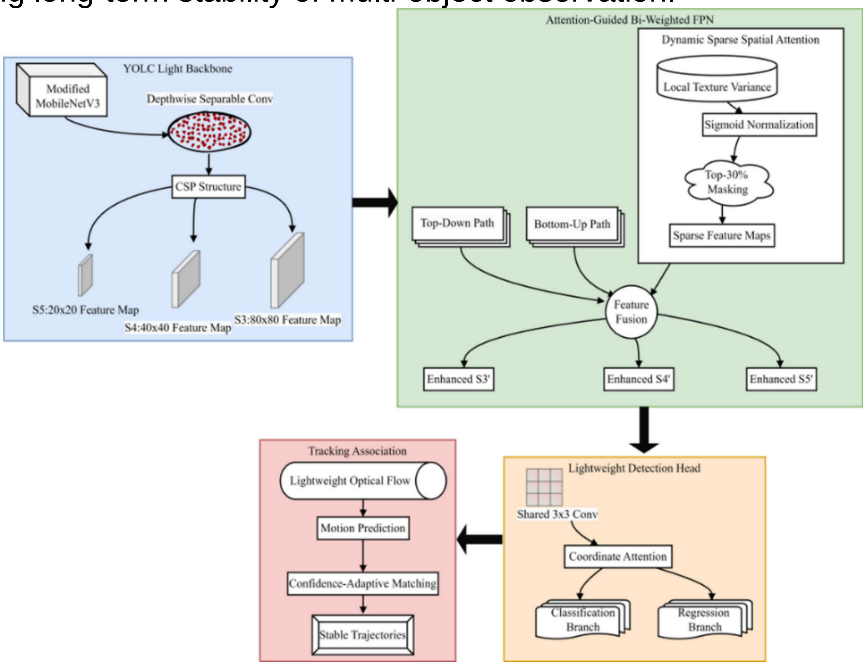

Fig.3. The general tracking association stage in Multiple Object Tracking.

Motion prediction and confidence-adaptive matching mechanisms, when combined with Siamese or metric learning methods, help generate long-term and stable trajectories of objects [4]. This stage is one of the main mechanisms determining the accuracy and reliability of the MOT system.

The objective assessment of the analyzed approaches was based on three main metrics widely used in the field of MOT: MOTA, IDF1, and FPS.

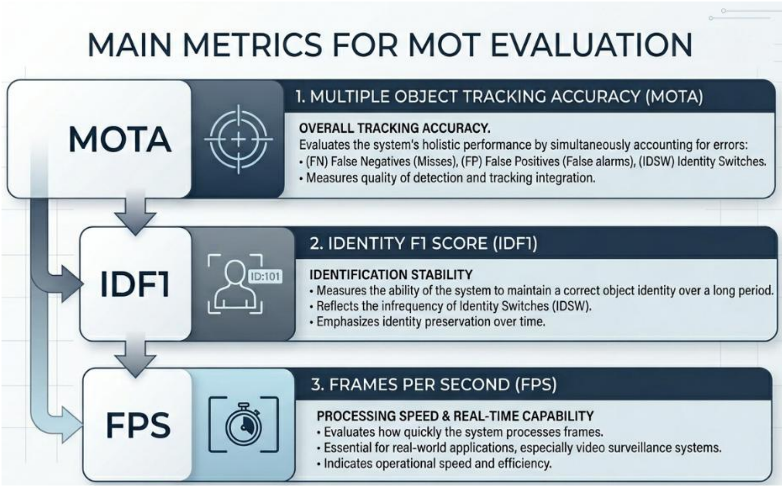

Fig. 4. Main metrics for MOT evaluation

Deep metric learning and Siamese methods can typically reduce ID switching by 20-40% through appearance-based features. Uncontrolled Re-ID approaches, on the other hand, are particularly useful when labeled data is scarce, making it easier to adapt to real-world conditions.

However, achieving high accuracy significantly increases computational complexity and resource requirements. As a result, maintaining balance in systems requiring real-time operation remains a pressing issue.

To solve this problem, a promising direction is the creation of a hybrid model by combining the occlusion estimation module (Liu et al., 2022) with SiamMOT or vMF-based approaches. Such a hybrid approach can provide high IDF1 and good real-time speed at the same time.

The SiamMOT method is clearly illustrated in the following diagram (Fig. 5). The Siamese neural network extracts feature vectors from two input images (the current frame and the previous frame) using the same weights (shared weights). The resulting vectors are evaluated using Multi-Layer Perceptron (MLP). As a result, detection and Re-ID tasks are performed simultaneously within the same network. This approach not only significantly reduces identifier switching (ID switch) but also increases computational efficiency.

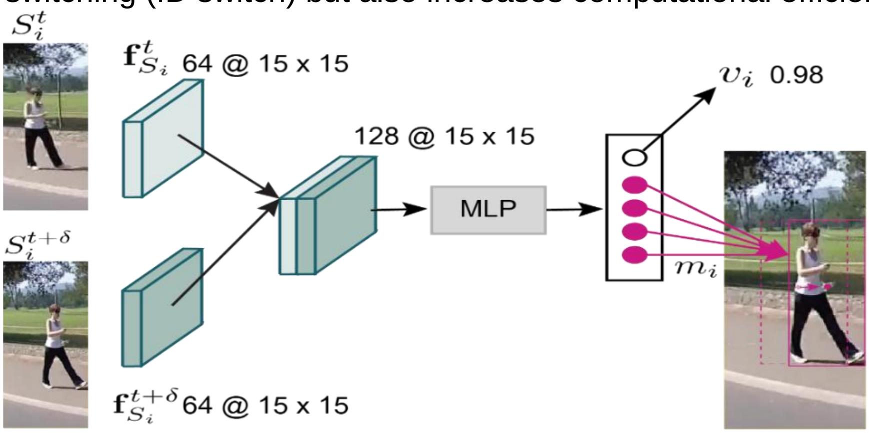

Fig.5. Siam MOT architecture.

Feature vectors are extracted from the two images via the Siamese network and the similarity is evaluated using MLP. This method combines detection and Re-ID to ensure tracking stability.

Feature maps and heatmaps play an important role in solving the occlusion problem. They help to identify objects that are missed by the detector or are left under an obstacle. Occlusion heatmaps generated by sigmoid activation predict an obstacle probability in each pixel within the range of [0,1], while an offset map allows for the correction of the object's exact location. This approach ensures continuous observation even under severe obstacle conditions.

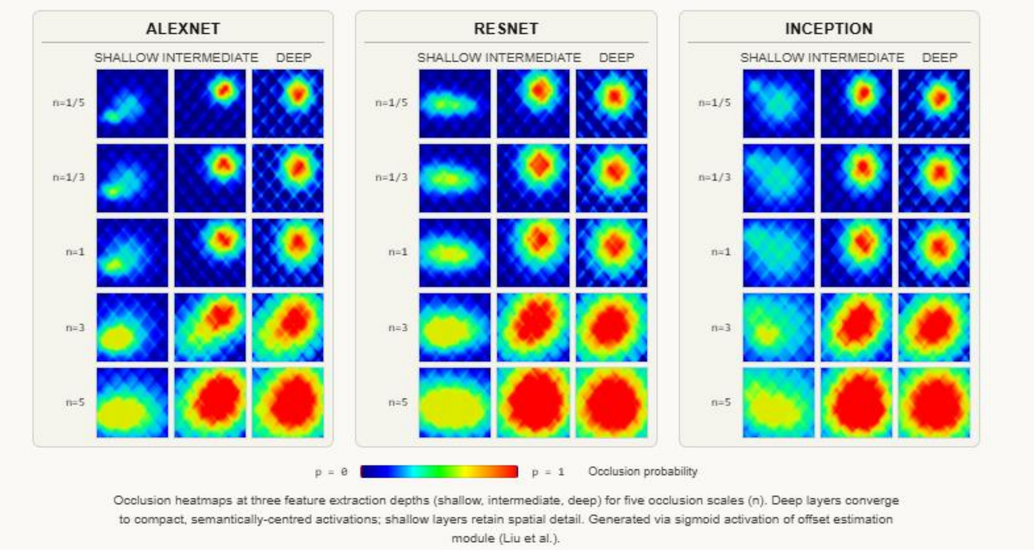

Fig. 6. Feature maps and occlusion heatmaps of various depths.

Occlusion centers (red-yellow areas) can be detected by sigmoid activation. This is suitable for visualizing the occlusion estimation module of Liu et al.

This figure shows three fundamental convolutional neural network architectures: AlexNet on the left, ResNet in the center, and the Inception model on the right.

Feature map and occlusion heatmap: Identify objects under a barrier in CNN architectures. Modern object recognition systems (AlexNet, ResNet, Inception) operate based on deep neural networks. In these networks, a feature map - a map of features isolated by each layer - and an occlusion heatmap - a heat map showing the probability of an obstruction - play an important role. CNN networks process information at three levels.

• Shallow layers only distinguish low-level features - lines, edges, and textures.

• Intermediate layers begin to recognize individual parts of an object, such as shapes and structures.

• Deep layers can understand the entire object as a whole and locate it accurately. Therefore, in deep layers, the active zone of the occlusion heatmap-the red-yellow zone-appears narrow and concentrated, while in surface layers, it appears wide and scattered.

The barrier level is expressed by the parameter n. At n=1/5, only a small part of the object is hidden, while at n=5, several objects overlap, and the barrier level is the highest. As the value of N increases, the active zone expands, and the model encounters greater uncertainty in determining the barrier center.

Occlusion heatmaps are generated by the sigmoid activation function. This function calculates the probability of an obstacle for each pixel in the interval

[0, 1]: 0 - no obstacle, 1 - complete obstacle. An offset map allows for the correction of an object's exact location. This approach ensures continuous observation even under severe obstacle conditions.

All three of these networks belong to the Convolutional Neural Network (CNN) family, but their design, depth, and operating principles differ significantly.

Fig. As shown in Figure 5, while AlexNet's feature maps are relatively simple, ResNet highlights deeper and more stable features due to skip connections. Inception, on the other hand, efficiently captures more data through filters of different sizes at the same time, showing rich and accurate results in occlusion heatmaps. The difference between these three networks affects the efficiency of Re-ID and MOT under conditions of heavy obstacles. While AlexNet's feature maps are relatively simple, ResNet highlights deeper and more stable features thanks to skip connections. Inception, on the other hand, efficiently captures more data through filters of different sizes at the same time, showing rich and accurate results in occlusion heatmaps. The difference between these three networks affects the efficiency of Re-ID and MOT under conditions of severe occlusion.

The methods used in the study have significant differences in functional capabilities and practical effectiveness. The Unsupervised Re-Identification approach is characterized by the fact that it does not require labeled data. This makes it suitable for use in large-scale data flows. At the same time, the main disadvantage of this method is that accuracy indicators are generally lower than those of supervised approaches, as the model lacks clear labels that guide the learning process.

The Occlusion Estimation method helps to restore the movement trajectory of lost objects in situations where a strong obstacle (occlusion) has occurred. This approach makes it easier to determine the continuous presence of an object in its temporary invisible state. However, since the method requires additional computing power as a separate module, it may affect the overall performance of the system.

The Siamese Deep Metric Learning (Thoreau) model is based on a combined analysis of appearance and motion features. These combined characteristics led to high results in benchmarks such as MOT16. Nevertheless, the detection boosting stage of the method must be managed very carefully, otherwise the number of incorrectly identified objects may increase.

The four different approaches proposed for Person Re-ID are distinguished by their stable performance in various camera points, lighting levels, and complex stage conditions [8]. A common disadvantage of these methods is that they require a large, clearly labeled database to function effectively.

The SiamMOT method increases operational efficiency due to the simultaneous execution of detection and re-identification processes and can provide real-time results at a speed of 17 FPS. However, the main limitation of the method is that it is primarily optimized for a single camera, and the stability of results in multi-camera surveillance systems may decrease.

We can offer a comparative analysis of the issues under consideration. Comparison of the main functional capabilities of Multi-Object Tracking and Re-ID methods.

Table 1. Analysis of computational resources and time complexity.

Method	Occlusion	ID Switch	Real-time	Labeled data	Multi-camera	Overall grade
Unsupervised Re-ID	✓	✓	✗	Not required	✗	Medium
Occlusion Estimation	✓✓	✓	✓	Requires	✓	Upper
Siamese DML (Thoreau)	✓	✓✓	✓	Requires	✓	Upper
Person Re-ID	✓	✓✓	✓	Requires	✓	Medium

Method	Occlusion	ID Switch	Real-time	Labeled data	Multi-camera	Overall grade
SiamMOT	✓✓	✓✓	✓	Requires	✗	Upper

✓✓ - resolves at a high level ✓ - partially resolves ✗ - doesn't decide

Analysis of computational resources and time complexity. This diagram presents a comparative analysis of the MOT methods considered in terms of speed, architecture, computational and time complexity, and device requirements. The diagram simultaneously demonstrates the computational efficiency and complexity of the MOT approaches. The high values for each axis indicate that the method is demanding for this parameter.

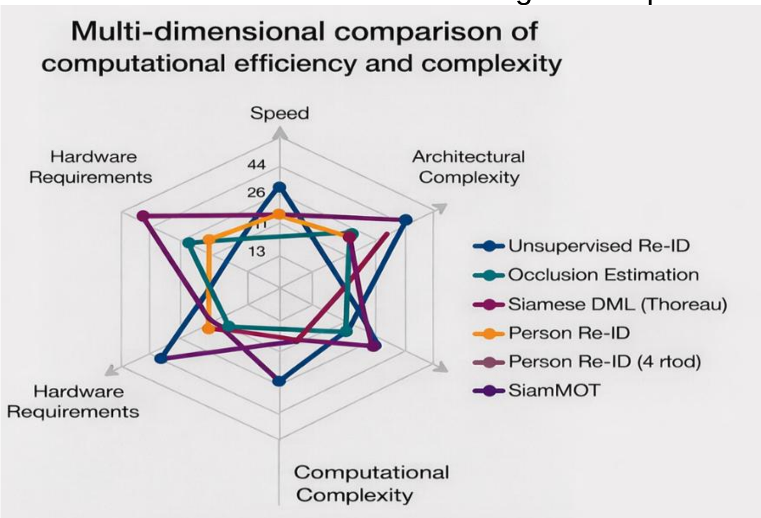

Fig.7 Analysis of computational resources and time complexity

The Siam MOT model is expected to demonstrate high efficiency in the agricultural sector, especially in the livestock sector. This is because multi-object monitoring technologies have not yet been fully implemented in this field, and existing monitoring systems rely mainly on traditional or semi-automated methods. The implementation of advanced tracking algorithms, such as Siam MOT, into livestock processes allows for precise real-time tracking of animal movements, reduces problems with identifier omissions or substitutions, and increases the reliability of digital monitoring.

Based on the results of this study, we can make the following proposals for the further development of the field of MOT.

1. It is recommended to develop a unified system that combines SiamMOT and the occlusion estimation module with unsupervised Re-ID to create a hybrid architecture. Currently, these three approaches have been studied separately, but their application in a combined architecture can further reduce ID switch errors and increase reliability under heavy barrier conditions.

CCTV cameras, mobile devices) are insufficient to function. Therefore, simplifying the model architecture, for example through knowledge distillation or pruning methods, expands the possibilities of practical application.

2. Creating a lightweight model The SiamMOT operates at a speed of 17 FPS - sufficient for real-time operation, but insufficient for operation on limited computing resources (drones, CCTV cameras, mobile devices). Therefore, simplifying the model architecture, for example through knowledge distillation or pruning methods, expands the possibilities of practical application.

3. Adaptation to a multi-camera system All the analyzed approaches were tested in a single-camera environment. In real-world conditions-an airport, a shopping mall, or road

surveillance several cameras work together. Solving the cross-camera Re-ID problem significantly increases the practical value of these systems.

4. Solving the problem of labeled data Supervised Re-ID approaches require a large amount of labeled data. Unsupervised Re-ID partially fixes this issue, but accuracy is still low. Combining GAN-based synthetic data generation with unsupervised learning can reduce dependence on labeled data.

5. Application in specialized fields the authors of SiamMOT themselves note, this system is effective for animal tracking in livestock farming. In this direction, the creation of specialized datasets and the adaptation of models for fields such as agriculture, medicine (instrumental observation in surgical videos), and sports analysis is a promising area of research.

For engineers and programmers, proposals for hybrid and lightweight architecture will guide the creation of practical systems. For subsequent researchers, the multi-chamber system and the GAN+unsupervised combination will serve as new research topics. Proposals for implementation in livestock, transport, and security systems for industry and business demonstrate the real-world value of the technology.

Conclusion

This study provides a comprehensive analysis of modern approaches aimed at improving the object re-identification process in multi-object tracking systems. To form a theoretical foundation, 200 scientific articles were studied, and advanced works on the Re-ID process, occlusion restoration mechanisms, and unsupervised learning opportunities were selected based on criteria of content relevance and practical effectiveness. As a result, five main scientific sources were selected that ensure stable tracking in MOT systems and aim to reduce ID switch events. During the analysis, three fundamental convolutional neural network architectures-AlexNet, ResNet, and Inception-were compared. AlexNet's low resource requirements, ResNet's superiority in forming deep embeddings through residual blocks, and Inception's ability to extract rich features with multidimensional convolution were identified as their strengths. At the same time, deep architectures such as ResNet and Inception have certain limitations when integrating into practical real-time systems due to their high computational and time complexity.

A comparative analysis of MOT methods in terms of speed, architectural complexity, computing power requirements, and adaptability to occlusion conditions showed that the SiamMOT model demonstrates the most optimal result, as it combines detection and identification processes in a single pipeline. This approach maintains the stability of identifiers even under occlusion conditions while operating at 17 FPS, significantly reducing ID switch cases. Although FairMOT and ByteTrack methods based on lightweight architectures provide high speeds, identification stability under occlusion conditions is not at the level of SiamMOT. This indicates that the problem of the global balance between accuracy and real-time speed remains relevant.

A comparative analysis of MOT methods in terms of speed, architectural complexity, computing power requirements, and adaptability to occlusion conditions showed that the SiamMOT model demonstrates the most optimal result, as it combines detection and identification processes in a single pipeline. This approach maintains the stability of identifiers even under occlusion conditions while operating at 17 FPS, significantly reducing ID switch cases. Although FairMOT and ByteTrack methods based on lightweight architectures provide high speeds, identification stability under occlusion conditions is not at the level of SiamMOT. This indicates that the problem of the global balance between accuracy and real-time speed remains relevant.

References:

[1]. Qiankun Liua, Dongdong Chenb, Qi Chua, Lu Yuanb, Bin Liua, Lei Zhangc, Nenghai Yu. Online Multi-Object Tracking with Unsupervised Re-Identification Learning and Occlusion Estimation. University of Science and Technology of China, Hefei, China. 4.01.2022.

[2]. Chenggui Sun, Li Bin Song. Product Re-identification System in Fully Automated Defect Detection.21.12.2021. current version: 1.0 Digital Object Identifier None

[3]. Michael Thoreau, Navinda Kottege. Improving Online Multiple Object tracking with Deep Metric Learning. 20.06.2018.

[4]. Zhangqiang Ming, Min Zhu, Xiangkun Wang, Jiamin Zhu, Junlong Cheng, Chengrui Gao, Yong Yang, Xiaoyong Wei. Deep learning-based person re-identification methods: A survey and outlook of recent works. 13.01.2022. College of Computer Science, Sichuan University, Chengdu, 610065, China.

[5]. Gioele Ciaparrone, Francisco Luque Sánchez, Siham Tabik, Luigi Troiano, Roberto Tagliaferri, Francisco Herrera. Deep learning in video multi-object tracking: a survey. 19.11.2019. Department of Management and Innovation Systems, University of Salerno, 84084 Fisciano (SA), Italy.

[6]. Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, Wenyu Liu. FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking. 19.10.2021. https://github.com/ifzhang/FairMOT

[7]. Bing Shuai Andrew, Berneshawi Xinyu Li, Davide Modolo, Joseph Tighe. SiamMOT: Siamese Multi-Object Tracking. https://github.com/amazon-research/siam-mot. 25.05.2021.

[8]. Babadjanov E.S., Toliev Kh.I. A Multi-Parameter Algorithm for Retagging Lost Animal RFID tags for PLF Platform // International Scientific and Practical Conference on «Actual Problems of Mathematical Modeling and Information Technology». Melville, New York. AIP Conference Proceedings. Vol.3147. 2024. –P. 050001-1–050001-8. https://doi.org/10.1063/5.0210338

[9]. E.S. Babadjanov, D.S. Serjanova. MTGAN arxitekturasi yordamida klinik vaqtli ma'lumotlarni generatsiya qilish usullari// Muhammad al-Xorazmiy avlodlari. Ilmiy-amaliy va axborottahliliy jurnal. ISSN-2181-9211. 3(33)/2025.