The CSI Journal on
Computer Science and Engineering
Vol. 20, No. 1, 2026
Pages 89-97
doi.org/10.22034/jcse.2025.558373.1065
Regular Paper

Online Multi-Object Tracking Using Convolutional Neural Networks and the Invasive Weed Optimization Algorithm

Amir AliAbadian[1]

[1]Department of Electrical and Biomedical Engineering, Shomal University

Abstract
Multi-object tracking (MOT) is a fundamental problem in computer vision with critical applications in areas such as video surveillance, human-computer interaction, autonomous driving, and video analytics. The main objective of MOT is to estimate the motion trajectories of multiple objects across sequential video frames while preserving their consistent identities throughout the sequence. MOT algorithms are generally categorized into two types: online methods, which process each frame sequentially and make tracking decisions in real-time, and offline methods, which process the entire video or segments of it as a batch to improve accuracy. In this study, we propose an online multi-object tracking method based on convolutional neural networks (CNNs). Unlike traditional approaches with fixed architectures, our method dynamically optimizes the number of hidden layers in the ANN using the Invasive Weed Optimization (IWO) algorithm, a nature-inspired metaheuristic optimization technique. This optimization aims to minimize the classification error, thereby enhancing the tracking performance by selecting a network architecture that is best suited to the complexity of the input data. The proposed system is evaluated using the VS-PETS 2009 benchmark dataset, a widely used dataset for evaluating object tracking algorithms. All simulations and model training are carried out in the MATLAB environment. The experimental results indicate that the proposed method achieves superior tracking accuracy and identity preservation performance compared to conventional tracking methods, demonstrating the effectiveness of combining ANNs with IWO in real-time multi-object tracking scenarios.
Keywords: Multi-Object Tracking, Video Analysis, Artificial Neural Network, Invasive Weed Optimization Algorithm.

1. Introduction

Multi-Object Tracking (MOT) is a key task in the field of computer vision and image processing, aiming to detect and track multiple moving objects throughout video sequences. Unlike single-object tracking, MOT requires maintaining the spatial locations and distinct identities of each object in every frame and correctly associating them across frames. This process not only identifies the objects' positions in the current frame but also ensures that each object is consistently tracked over time [1-4]. Generally, MOT methods are divided into two main categories: online and offline. In online methods, objects are tracked frame-by-frame in real-time, making decisions based only on past and current data, which is crucial for real-time applications. In contrast, offline methods process the entire video or segments of it as a batch, allowing the algorithm to better analyze complex associations between objects, albeit without real-time capability.

The importance of MOT is evident in numerous intelligent systems and everyday applications. In surveillance, security systems must accurately track people, vehicles, and moving objects in crowded and complex environments to promptly detect and respond to incidents. In human-computer interaction, precise multi-object tracking enables understanding of human behavior and gesture recognition. Moreover, in autonomous vehicles, the ability to continuously detect and track surrounding objects such as cars, pedestrians, and cyclists is fundamental for safe decision-making and accident prevention [5-6].

The applications of MOT are extensive and go beyond typical surveillance and autonomous driving. For example, in sports video analysis, MOT facilitates tracking players and the ball to improve performance evaluation and strategy. In robotics, multi-object tracking helps robots better understand their environment and react appropriately to moving objects. In biological sciences, MOT is used to track the movement paths of cells or particles in microscopic images, aiding in the understanding of biological processes [7-8]. In all these applications, high accuracy and speed of tracking are critical for system success.

Despite significant advances, MOT faces numerous challenges that affect its performance. One of the major challenges is occlusion, where one or more objects are temporarily hidden by others, complicating continuous tracking. Additionally, physical overlap of objects, sudden changes in object appearance due to varying viewpoints, lighting conditions, or scale changes, pose significant difficulties. In crowded scenes with many simultaneously moving objects, correctly maintaining object identities (data association) is especially challenging, and errors here can cause identity switches [9-12].

In this context, Artificial Neural Networks (ANNs) have emerged as a powerful tool to overcome the challenges of multi-object tracking [13-16]. Deep models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs and LSTMs) can extract complex semantic features from images, which are essential for better object detection and differentiation. Moreover, neural networks can learn patterns of object motion and appearance, improving the prediction of their future positions, which helps maintain identity consistency and tracking accuracy [17-21].

Furthermore, combining neural networks with optimization algorithms and advanced learning methods, such as the Invasive Weed Optimization (IWO) algorithm, enables optimizing the network architecture and parameters to reduce classification errors and enhance tracking performance. These hybrid approaches improve accuracy while reducing computational complexity, which is critical for real-time implementation.

This paper is organized as follows: Section 2 provides a comprehensive review of existing multi-object tracking methods. Section 3 details the proposed online multi-object tracking framework, including the integration of artificial neural networks with the Invasive Weed Optimization algorithm, and explains its theoretical background and implementation. Section 4 presents the simulation results conducted on benchmark datasets, offering a comparative analysis of the proposed method against state-of-the-art techniques. Finally, Section 5 concludes the presented paper.

2. Existing Methods

Reference [1] introduces a detection-based multi-object tracking framework that utilizes a two-stage data association strategy to achieve high accuracy and consistent tracking. In the first stage, reliable short-term tracks are generated solely based on spatial information. The second stage performs global-level linking of these short-term tracks to maintain continuity over longer periods. The proposed tracker was tested on the ABQ dataset, where it showed promising performance compared to existing approaches. The method presented in [2] is built upon a robust detection pipeline comprising multiple object detectors. Its performance was evaluated using the MOT16 benchmark metrics. The proposed approach demonstrated superior speed compared to other online trackers on the challenging MOT16 dataset, while still preserving a satisfactory level of accuracy. Object tracking faces significant challenges arising from factors like noise, similarity between objects, and low image resolution. In [3], an entropy minimization-based tracking method is proposed, which effectively handles diverse data sequences. The study demonstrates that the approach maintains high accuracy even in difficult scenarios involving illumination variations, occlusions, and rapid camera movements—conditions commonly encountered in videos captured by moving cameras. The multi-object tracking system built on this technique was implemented and extensively evaluated on several benchmark datasets, confirming its robust performance.

In [4], a bidirectional network for MOT is proposed, which utilizes both forward and backward tracking for joint object tracking. Additionally, a near-online tracking model based on this network is introduced to enhance performance by classifying tracks and performing forward and backward tracking in parallel. Experimental results on the challenging MOT benchmarks demonstrate that the proposed method outperforms existing approaches. One of the key tasks in multi-object tracking is data association. The study in [5] discusses data association in multi-object tracking and complements it with a two-layer network flow approach. The association probability is computed based on positional distance and feature similarity between objects. The dataset used in this research is 2DMOT2015. Tracking can easily fail when objects become occluded. In [6], a method is proposed to address this issue. To balance accuracy and speed, a three-stage data association process is introduced. In the first stage, trajectories and targets are matched directly. The second stage matches trajectories and targets based on smaller overlapping regions. Finally, in the third stage, the Hungarian algorithm is employed to associate the remaining unmatched trajectories and targets.

In the field of multi-object tracking, several studies have focused on using CNNs to extract appearance features of target objects. These appearance features include patterns, colors, and textures that help to better distinguish and accurately identify objects, playing a crucial role in maintaining object identities throughout video sequences. The use of such features significantly improves tracking accuracy and enhances the ability to differentiate similar objects. However, this advantage comes at the cost of high computational complexity, as extracting and processing appearance features with CNNs requires substantial computational power and can considerably increase processing time. This increase in processing time poses a major challenge for deploying these methods in real-time tracking systems, where processing speed is as important as accuracy. Therefore, in study [7], a simple and efficient method was proposed to improve processing speed by removing redundant and unnecessary appearance features. This technique reduces computational load, accelerates the algorithm’s performance, while maintaining tracking accuracy. The results presented in this study demonstrate that the proposed method achieves an optimal balance between high accuracy and suitable speed for real-time processing. In other words, this approach enables the tracking algorithm to be not only accurate but also practical and applicable in real-world scenarios. Consequently, this technique can be considered an effective solution for developing efficient and practical multi-object tracking systems. The correlation matching algorithm is employed for multi-object tracking in [8]. Experimental results demonstrate that the improved algorithm outperforms standard methods in handling target occlusion, scale variations, background interference, and re-identification of lost targets. This paper effectively extends the enhanced correlation filter algorithm to multi-target tracking, enabling better utilization on multi-core platforms.

3. Proposed Method

Since the proposed method is based on Convolutional Neural Networks and the Invasive Weed Optimization algorithm, first these techniques are reviewed, and then the proposed method is presented based on them.

3.1. Convolutional Neural Networks

CNNs are among the most widely used and successful deep learning architectures in the field of computer vision. Due to their unique structure and ability to extract spatial and structural features from images, they have been extensively applied to various tasks, including MOT. Unlike traditional neural networks, CNNs utilize convolutional layers that operate locally by applying learnable filters on small regions of the input image. This capability allows the network to automatically extract important patterns and features such as edges, textures, and shapes [17-21].

The basic structure of a CNN consists of several types of layers arranged in a specific order. The first layer is the input layer, which receives the raw image or data. Next are the convolutional layers that apply various filters to generate feature maps. Afterward, pooling layers are used to reduce the spatial dimensions and focus on the most important features. Finally, fully connected layers convert the extracted features into the final classification or regression outputs. Each of these layers plays a critical role in preserving spatial structure and extracting multi-level features.

In the context of Multi-Object Tracking, CNNs offer significant advantages. First, CNNs have the ability to extract complex and meaningful appearance features of objects, which helps reliably distinguish similar objects. This is crucial for maintaining the identity of each object throughout the video sequence. Second, convolutional networks can extract features that are more robust to variations in illumination, viewpoint, scale, and background details, which are essential for real-world tracking scenarios. Moreover, CNNs have the capacity for automatic learning and continuous improvement through large-scale training data, allowing them to adapt to diverse and challenging environments. Additionally, with advancements in hardware and software, these networks can be efficiently executed in real-time, which is a fundamental requirement for multi-object tracking systems. CNNs also improve detection and differentiation of objects under occlusion and overlap conditions, which are among the biggest challenges in MOT.

Long Short-Term Memory (LSTM) and CNNs are two distinct types of neural networks that are often combined to achieve more efficient and powerful performance [22-23]. In applications requiring the integration of spatial and temporal features, incorporating LSTM layers into CNNs can significantly enhance results. The role of the LSTM layer within a CNN is twofold: first, while CNNs excel at extracting spatial features and recognizing patterns in two- or three-dimensional data such as images or videos, LSTMs are known for managing sequential and temporal information. Thus, adding LSTM to CNN allows the network to capture temporal patterns like time-dependent changes in videos. Second, CNNs initially extract spatial features from the input data, which are then passed to the LSTM layer to understand temporal relationships or sequential dependencies. This combination is particularly effective in video recognition, time series prediction, and analysis of sequential data such as EEG signals or sensor outputs.

Finally, combining CNNs with other optimization and machine learning algorithms, such as the Invasive Weed Optimization algorithm, can further enhance the accuracy, speed, and stability of tracking systems. These combinations help in selecting optimal parameters and network structures, ultimately optimizing the overall system performance and yielding highly successful results in practical and industrial MOT applications.

3.2. Invasive Weed Optimization

The IWO algorithm is a nature-inspired metaheuristic optimization technique that mimics the natural growth and reproduction behavior of invasive weeds in an ecosystem. It has gained popularity due to its simplicity, flexibility, and effectiveness in solving a wide range of complex optimization problems across various engineering, science, and industrial fields [24-27]. The algorithm leverages biological concepts such as reproduction, dispersal, competition, and survival of the fittest to iteratively improve candidate solutions within the search space. The first step in the IWO algorithm is population initialization, where a number of initial solutions, called weeds, are randomly distributed throughout the problem's search space. Each weed represents a potential solution, encoded as a vector of parameters relevant to the optimization problem. This diversity in the initial population ensures a broad exploration of the search space, reducing the likelihood of premature convergence to local optima. Following initialization, the reproduction phase begins. In this phase, each weed generates a certain number of seeds based on its fitness value, which measures how well the solution solves the optimization problem. Weeds with higher fitness produce more seeds, thereby increasing the chances that their genetic material (solution features) propagate through the population. This mechanism mimics natural reproduction where fitter plants spread more offspring.

Once seeds are produced, they undergo seed dispersal in the search space. The dispersion is modeled using a Gaussian distribution, allowing seeds to spread around the parent weed's position with a certain variance. Importantly, this variance decreases over successive iterations, enabling a gradual shift from exploration (wide search) to exploitation (fine-tuning near promising solutions). As a result, seeds gradually converge towards the best areas of the search space. The algorithm also incorporates a natural selection process, often called competitive exclusion, to maintain a manageable population size. When the number of weeds exceeds a predefined maximum, the weakest weeds—those with lower fitness—are removed from the population. This step ensures that computational resources are focused on the most promising solutions and helps maintain diversity without overcrowding.

Finally, these steps—reproduction, dispersal, and selection—are repeated iteratively in a loop. The algorithm continues until a stopping condition is met, such as reaching a maximum number of iterations or achieving a satisfactory solution quality. Over successive iterations, the population evolves and converges towards optimal or near-optimal solutions. Due to its biologically inspired mechanisms and adaptive search strategies, the IWO algorithm is especially effective for nonlinear, multimodal, and high-dimensional optimization problems, making it a valuable tool in areas such as control systems, machine learning, and engineering design optimization.

3.3. Proposed MOT Method

The overall flowchart of the proposed MOT method is illustrated in Figure 1, which outlines the step-by-step process of the algorithm in a structured manner. This method integrates deep learning using a CNN with an evolutionary optimization technique, namely the IWO algorithm. The objective is to enhance performance in the task of MOT. In the proposed CNN–IWO framework, the CNN outputs are primarily used as feature embeddings for multi-object tracking. Each detected object is passed through the convolutional and LSTM layers to extract a feature vector that captures both spatial and temporal characteristics of the object. These feature vectors are then used for re-identification, i.e., associating objects across consecutive frames to maintain consistent identities.

While the fully connected and softmax layers provide classification outputs, the feature embeddings obtained before the fully connected layer serve as the main descriptors for matching objects between frames. The re-identification process is performed by comparing these embeddings to previously tracked objects using Euclidean distance, enabling the system to update object trajectories and preserve identities throughout the video.

Before applying the proposed CNN–IWO framework for multi-object tracking, the input video frames are preprocessed to enhance quality and facilitate robust feature extraction. The preprocessing steps include resizing all frames to a fixed resolution of 640×480 pixels and normalizing pixel intensity values to the range [0, 1]. Optional Gaussian smoothing is applied to reduce noise when necessary.

For object detection, a frame-by-frame detection approach is employed. In this study, a simple detection method such as background subtraction is used to identify moving objects in each frame. Background subtraction compares the current frame with a reference background model to locate regions of change, which correspond to objects of interest. The detected object bounding boxes are then provided as the initial input to the CNN–IWO tracking framework, which maintains object identities and estimates their trajectories across the video sequence.

The key stages of the proposed framework are described as follows:

1. Definition of the Initial Neural Network Architecture: The first step involves defining the overall architecture of the neural network. This includes specifying the types of layers, their order, activation functions, loss function, learning rate, and other network parameters. This base architecture serves as the skeleton of the learning model, upon which further optimization will be performed.
2. Identification of Design Parameters to Be Optimized: One of the critical design parameters in the CNN is the number of convolutional layers (Convolution2DLayer). Choosing the optimal number of these layers has a direct impact on the network's feature extraction capability and generalization power. In this method, the number of convolutional layers is treated as a design variable, and the goal is to find its optimal value such that the classification error is minimized.
3. Application of the IWO Algorithm for Optimization: The IWO algorithm, a population-based metaheuristic method, is used to determine the optimal number of convolutional layers. Through the generation and dispersal of seeds (candidate solutions) across the search space, the algorithm gradually converges towards more promising regions and optimizes the network's structure.
4. Classification Error Evaluation for Each Solution: For each solution generated by the IWO algorithm (e.g., a specific number of layers), a CNN model is constructed and trained accordingly. The performance of the model is then evaluated using classification error on validation data. This error value is fed back into the IWO algorithm as a fitness measure to guide the search process.
5. Stopping Criterion Check: After evaluating each generation of candidate solutions, the algorithm checks whether a stopping condition is satisfied. This may be a maximum number of iterations or achieving a classification error below a predefined threshold. If the stopping condition is not met, IWO generates a new set of solutions, and the optimization continues.
6. Applying the Optimized Structure to the Final CNN: Once the optimization process concludes and the best value for the design parameter is identified, this value is used to configure the final CNN structure. The finalized CNN is then trained and utilized for multi-object tracking in the target application.

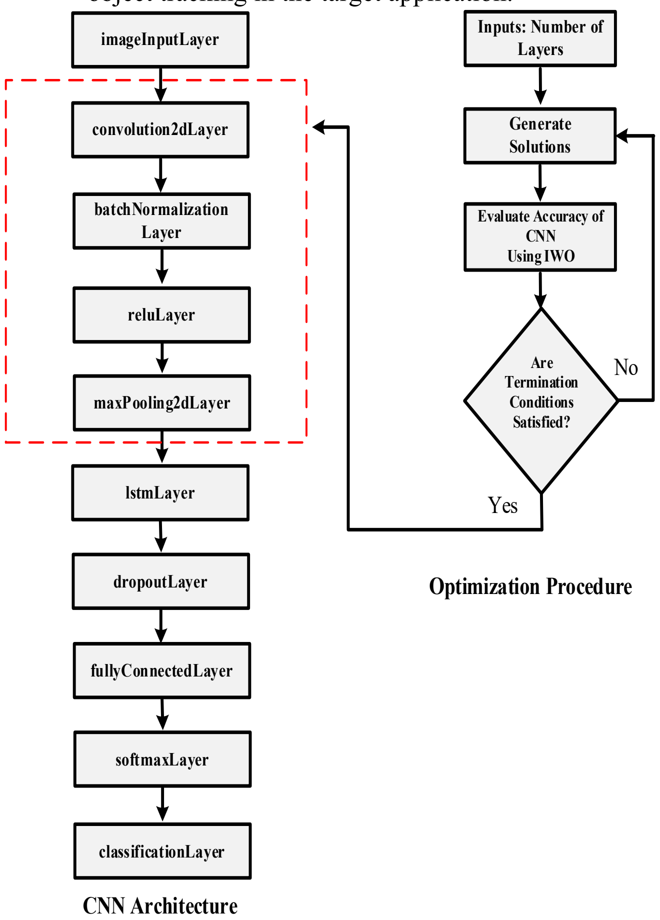

Figure 1. Flowchart of the proposed MOT method.

4. Simulation Results

The proposed and benchmark methods were implemented in MATLAB (MathWorks Inc.) and executed on a 64-bit Windows system equipped with an Intel® Core™ i5-4460 processor (3.20 GHz) and 32 GB of RAM. Table 1 presents the detailed configuration of the parameters employed for the IWO algorithm. These parameters were selected based on a balance between computational efficiency and solution quality. The tuning process was guided by insights from prior studies and a series of preliminary simulations to ensure a desirable convergence rate while mitigating the risk of premature convergence. Proper adjustment of these parameters enables the algorithm to effectively explore the search space and intensify the search around promising regions, ultimately leading to robust and high-quality solutions.

The baseline CNN architecture used in this study is defined with fixed structural parameters, while the number of convolutional layers is treated as the design variable optimized by the IWO algorithm. The architectural details are as follows:

- Convolutional filters: All convolutional layers use a kernel size of 3×3.
- Stride: A stride of 1 is applied in the convolutional layers, whereas the max-pooling layers use a stride of 2.
- Activation function: ReLU is employed in all hidden layers to introduce nonlinearity.
- Pooling: Max-pooling layers with a 2×2 window are inserted after selected convolutional layers for spatial downsampling.
- Fully connected layer: The network includes one dense layer with 128 neurons before the output layer.
- Output layer: A Softmax activation function is used for multi-class prediction.

This configuration serves as the foundation for the optimization process and ensures consistency across all candidate architectures evaluated by the IWO algorithm.

The training configuration of the CNN model is established using a set of standardized and widely adopted hyperparameters to ensure stable and efficient convergence. The cross-entropy loss function is utilized as the objective function, as it provides an effective measure of the discrepancy between predicted probabilities and the ground-truth class labels in multi-class classification tasks. Parameter updates are carried out using the Adam optimizer, which combines the advantages of momentum and adaptive learning rates, making it particularly suitable for handling noisy gradients and non-stationary training conditions.

A batch size of 32 is selected to achieve a practical balance between training stability, memory efficiency, and the representativeness of gradient estimates. The initial learning rate is set to 0.001, and a step-wise learning rate schedule is employed during training. Under this schedule, the learning rate is decreased by a factor of 0.1 every 10 epochs, which helps prevent premature convergence and facilitates more refined updates as the model approaches the optimal solution. This training configuration ensures robust convergence across all candidate architectures generated during the IWO optimization process.

Table 1. Parameters of the IWO

Parameters	Values
N_{weed}	50
MaxIt	5
P_{max}	50
S_{max}	10
S_{min}	1
$\delta_{initial}$	0.3
δ_{final}	0.001
pow	3

The MOT algorithm is evaluated using several datasets. The dataset used is VS-PETS 2009, which provides responses similar to those in [12], where sequences S2.L1 are utilized. The evaluation metrics include MT (Mostly Tracked), PT (Partially Tracked), ML (Mostly Lost), Frag (Fragmentation), and IDS (ID Switches). Examples of frames from the S2.L1 sequences are presented in Figures 2 to 5.

Furthermore, Figures 6 to 9 illustrate feature representations extracted by the three convolutional layers of the neural network, each capturing different levels of abstraction from the input images. As the network depth increases, these features progressively evolve from capturing simple low-level details to complex, high-level patterns and semantics. This hierarchical architecture enables the model to more accurately distinguish critical image components, enhancing its ability to handle challenges and noise present in real-world scenarios during object detection and tracking tasks. Such capabilities are crucial for improving the overall performance and reliability of computer vision systems.

In the process of optimizing the number of hidden layers in the neural network using the IWO algorithm (Figure 10), the primary objective is to minimize the cost function. During the initial iterations, variations in the number of layers lead to significant reductions in the cost, resulting in rapid improvements in the model's performance. However, after a few early iterations—specifically after the third iteration in this study—the changes in cost become negligible, and the optimization process approaches convergence. This behavior indicates that the algorithm has reached, or is close to, an optimal configuration where the network architecture achieves satisfactory performance in terms of the cost function. Such a trend is expected and confirms that the IWO algorithm effectively determines the optimal number of hidden layers, beyond which further optimization does not yield meaningful improvements.

Figure 2. The first frame from the S2.L1 sequence.

Figure 3. The second frame from the S2.L1 sequence.

Figure 4. The third frame from the S2.L1 sequence.

Figure 5. The fourth frame from the S2.L1 sequence.

Figure 6. Feature maps extracted from the convolutional layer corresponding to the Figure 2.

Figure 7. Feature maps extracted from the convolutional layer corresponding to the Figure 3.

Figure 8. Feature maps extracted from the convolutional layer corresponding to the Figure 4.

Figure 9. Feature maps extracted from the convolutional layer corresponding to the Figure 5.

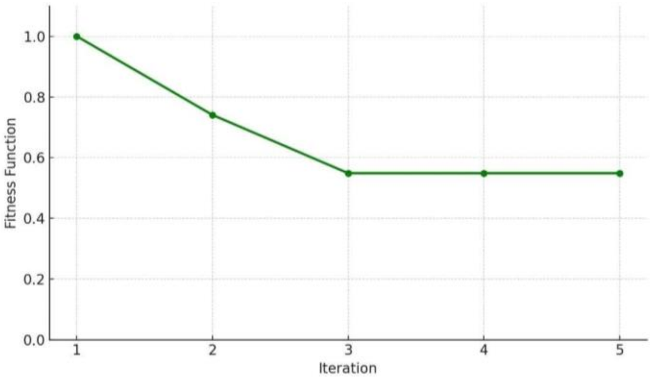

Figure 10. Fitness Function vs Iteration.

Table 2 provides a comprehensive comparison of the proposed Multiple Object Tracking (MOT) algorithm against existing methods. In this table, methods marked with an asterisk (*) are classified as online algorithms, whereas the remaining approaches are considered batch methods. For the VS-PETS 2009 dataset, experimental results demonstrate that the proposed method achieves superior performance compared to many batch-based algorithms across several key metrics. This highlights the proposed algorithm's capability for efficient and effective data processing, particularly in scenarios where tracking accuracy and reliability are critical. Such advancements have significant implications for practical applications in fields such as video surveillance, security, and video analytics.

Table 2. Compares the performance on the PETS S2.L1 dataset.

Method	MT	PT	ML	Frag	IDS
[10]	91.3%	4.4%	4.3%	6	11
[9]	78.9%	21.1%	0%	23	1
[12]	89.5%	10.5%	0%	9	0
[11] *	94.7%	5.3%	0%	7	0
Proposed MOT*	97.6%	2.4%	0%	6	0

5. Conclusions

Deep learning has revolutionized the field of MOT due to its superior ability to process complex and high-dimensional data, as well as its capacity for automatic feature learning. This advancement has led to significant improvements in various applications, including autonomous driving, video surveillance, and video analytics, where accurate and robust tracking of multiple objects simultaneously is crucial.

One of the key advantages of deep learning in MOT is its enhanced object detection capabilities. By integrating sophisticated object detection frameworks with tracking algorithms, modern deep learning models can reliably identify and track objects even in challenging scenarios. These scenarios may involve significant changes in object appearance, partial occlusions, or temporary disappearance from the field of view. In this research, we propose a novel deep learning-based MOT method that leverages an adaptive approach to optimize the number of convolutional layers within the network. This optimization is achieved through the application of the IWO algorithm, which dynamically selects the architecture parameters to balance model complexity and performance. Extensive simulation results validate the proposed method's effectiveness, demonstrating superior tracking accuracy and robustness compared to existing approaches. The proposed framework not only enhances the precision of object localization and identification but also improves the overall efficiency of the tracking system, making it highly suitable for real-world applications with complex and dynamic environments.

References

[1] Al-Shakarji, N. M., Bunyak, F., Seetharaman, G., & Palaniappan, K. (2018). Robust multi-object tracking for wide area motion imagery. *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE. https://doi.org/10.1109/AIPR.2018.8707377

[2] Cobos, R., Hernandez, J., & Abad, A. G. (2019). A fast multi-object tracking system using an object detector ensemble. *2019 IEEE Colombian Conference on Applications in Computational Intelligence (ColCACI)*. IEEE. https://doi.org/10.1109/ColCACI.2019.8781972

[3] Jerome, C. R., Nayak, S. D., Malagi, V. P., & Rangarajan, K. (2018). Entropy minimization based multi object tracking. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE. https://doi.org/10.1109/ICCONS.2018.8662974

[4] Zhong, X., Tan, M., Ruan, W., Huang, W., Xie, L., & Yuan, J. (2020). Dual-direction perception and collaboration network for near-online multi-object tracking. *2020 IEEE International Conference on Image Processing (ICIP)*. IEEE. https://doi.org/10.1109/ICIP40778.2020.9190645

[5] Abdillah, B., Jati, G., Jatmiko, W., & Nurhadiyatna, A. (2018). Two layer network flow for fast data association on multi object tracking. *2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE. https://doi.org/10.1109/ICACSIS.2018.8618161

[6] He, S., & Wang, P. (2019). Online multi-object tracking with three-stage data association. *2019 IEEE International Conference on Computational Electromagnetics (ICCEM)*. IEEE. https://doi.org/10.1109/CompEM.2019.8778960

[7] Yoon, Y.-C., Song, Y.-M., Yoon, K., & Jeon, M. (2018). Online multi-object tracking using selective deep appearance matching. *2018 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*. IEEE. https://doi.org/10.1109/ICCE-Asia.2018.8552105

[8] Wu, Q., Xue, W., Zheng, X., & He, W. (2020). Online multi-object tracking based on improved kernel correlation filter. *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*. IEEE. https://doi.org/10.1109/ICAICA50127.2020.9182383

[9] Kuo, C.-H., & Nevatia, R. (2011). How does person identity recognition help multi-person tracking? *IEEE Conference on Computer Vision and Pattern Recognition*, 1217–1224.

[10] Milan, A., Roth, S., & Schindler, K. (2014). Continuous energy minimization for multitarget tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(1), 58–72.

[11] Xu, L., Li, W., Wu, H., & Li, Q. (2016). Online multi-object tracking based on global and local features. *IEEE Visual Communications and Image Processing (VCIP)*, 1–4.

[12] Yang, B., & Nevatia, R. (2012). Multi-target tracking by online learning of non-linear motion patterns and robust appearance models. *IEEE Conference on Computer Vision and Pattern Recognition*, 1918–1925.

[13] Zhao, D., Fu, H., Xiao, L., Wu, T., & Dai, B. (2018). Multi-object tracking with correlation filter for autonomous vehicle. *Sensors, 18*(7). https://doi.org/10.3390/s18072026

[14] Liu, Q., Liu, B., Wu, Y., Li, W., & Yu, N. (2019). Real-time online multi-object tracking in compressed domain. *IEEE Access, 7*, 76489–76499. https://doi.org/10.1109/ACCESS.2019.2922181

[15] Li, X., Jiang, K., Jia, S., Zhang, X., & Sun, Y. (2018). Visual tracking based on convolutional neural network. In *2018 Chinese Automation Congress (CAC)* (pp. 4061–4066). IEEE.

[16] Kong, X., Zhang, B., Yue, L., & Xiao, Z. (2018). Attentional convolutional neural networks for object tracking. In *2018 Integrated Communications, Navigation, Surveillance Conference (ICNS)* (pp. 5B1-1–5B1-11). IEEE.

[17] Cho, M., & Kim, E. (2025). 3D LiDAR multi-object tracking using multi positive contrastive learning and deep reinforcement learning. *IEEE Access, 13*, 12447–12457. https://doi.org/10.1109/ACCESS.2024.3521334

[18] Luo, J., & Zhang, Y. (2025). Research on multi-target tracking method based on deep learning. In *2025 5th International Conference on Sensors and Information Technology* (pp. 464–467). IEEE. https://doi.org/10.1109/ICSI64877.2025.11009782

[19] Cheng, L., Sengupta, A., & Cao, S. (2024). Deep learning-based robust multi-object tracking via fusion of mmWave radar and camera sensors. *IEEE Transactions on Intelligent Transportation Systems, 25*(11), 17218–17233. https://doi.org/10.1109/TITS.2024.3421339

[20] Farkhodov, K., Lee, S.-H., Platos, J., & Kwon, K.-R. (2023). Deep reinforcement learning TF-agent-based object tracking with virtual autonomous drone in a game engine. *IEEE Access, 11*, 124129–124138. https://doi.org/10.1109/ACCESS.2023.3325062

[21] Jiao, L., Wang, D., Bai, Y., Chen, P., & Liu, F. (2023). Deep learning in visual tracking: A review. *IEEE Transactions on Neural Networks and Learning Systems, 34*(9), 5497–5516. https://doi.org/10.1109/TNNLS.2021.3136907

[22] Nijhum, I. R., Hossain, M. Y., Akinola, T., Rakib, M. M. H., & Dev, S. (2024). A hybrid CNN-LSTM framework and infrared image processing for solar irradiance forecasting. In *2024 17th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 1–6). IEEE. https://doi.org/10.1109/CISP-BMEI64163.2024.10906220

[23] Zheng, R., et al. (2022). Automatic liver tumor segmentation on dynamic contrast enhanced MRI using 4D information: Deep learning model based on 3D convolution and convolutional LSTM. *IEEE Transactions on Medical Imaging, 41*(10), 2965–2976. https://doi.org/10.1109/TMI.2022.3175461

[24] Mehrabian, A. R., & Lucas, C. (2006). A novel numerical optimization algorithm inspired from weed colonization. *Ecological Informatics, 1*(4), 355–366.

[25] Luo, X., Li, M., Feng, N., & Chen, F. (2017). Optimal design on customized bundling strategy of information goods for customers with two-dimensional heterogeneity. *Asia-Pacific Journal of Operational Research, 34*(2), 1750007. https://doi.org/10.1142/S0217595917500075

[26] Angeletti, P., Berretti, L., Maddio, S., Pelosi, G., Selleri, S., & Toso, G. (2022). Phase-only synthesis for large planar arrays via Zernike polynomials and invasive weed optimization. *IEEE Transactions on Antennas and Propagation, 70*(3), 1954–1964. https://doi.org/10.1109/TAP.2021.3119113

[27] Saxena, A., Kumar, R., Singh, J., Singh, G. K., Kumar, V. S., & Pandey, J. P. (2022). An invasive weed optimization for sensorless control of grid integrated wind driven doubly fed induction generator. *IEEE Access, 10*, 109082–109096. https://doi.org/10.1109/ACCESS.2022.3213982

Amir AliAbadian Department of Electrical and Biomedical Engineering, Shomal University.

Email: a.aliabadian@shomal.ac.ir

Paper Handling Data:
Corresponding author: **Amir AliAbadian**
Affiliation of the corresponding author: Department of Electrical and Biomedical Engineering, Shomal University.