Enhanced Multi-Object Detection via the Integration of PSO, Kalman Filtering, and CNN Compressive Sensing

S. V. Suresh Babu Matla[1], Dr.S. Ravi[2*], Muralikrishna Puttagunta[3]
Research Scholar, Department of Computer Science-School of Engineering and Technology, Pondicherry University, India[1]
Professor, Department of Computer Science-School of Engineering and Technology, Pondicherry University, India[2]
Assistant professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, 522502, Andhra Pradesh, India[3]

***Abstract*—Many inventive techniques have been created in the field of machine vision to solve the challenging challenge of detecting and tracking one or more objects in the face of challenging conditions, such as obstacles, object motion, changes in light, shaking, and rotations. This research article provides a novel method that combines Convolutional Neural Networks (CNNs), Compressive Sensing, Kalman Filtering, and Particle Swarm Optimization (PSO) to address the challenges of multi-object tracking under dynamic conditions. Initially, a CNN-based object classification and identification system is demonstrated, which efficiently locates objects in video frames. Subsequently, in order to produce precise representations of object appearances, utilizing compressive sensing techniques. The Kalman Filter ensures adaptability to irregular observations, eliminates erroneous data, and reduces uncertainty. PSO enhances tracking efficiency by optimizing forecast precision. When combined, these techniques provide robust tracking even in the presence of complex movement patterns, occlusions, and visual disparities. The efficiency of this strategy is demonstrated by an empirical investigation that produces a remarkable tracking accuracy of 98%, which is 3.15% greater than other methods across a range of challenging settings. This technique has been compared to various existing approaches, including the Clustering Method, YOLOV4 DNN Model, and YOLOV3 Model, and its deployment is made easier with Python software. This hybrid technique, which addresses the limitations of separate approaches and offers a holistic approach to multi-object monitoring, has potential applications in surveillance, robotics, and autonomous systems.**

Keywords—Multi-Object tracking; object detection; convolutional neural networks; kalman filtering; particle swarm optimization

I. Introduction

Computer vision is a dynamic investigation area, which widens its perceptual field across a range of purposes like traffic and navigation systems, motion detection, speech and voice recognition and other applications. In which Object detection by means of a single aspect of the target will not provide a favorable resolution, as the object manifestation may altered depending on light effects, object position shifts, and blurring effects, occlusion and so on. As a result, multi-scale object detection was planned, which delivers further healthy resolution than single-degree object detection [1]. Quite a lot of discriminative appearance Random projections have been expected to venture high dimensions properties like features to a low dimensional space. This compressive sensing is self-sufficient in terms of data, as well as non-adaptive naïve Bayes classifier is used to differentiate among optimistic and undesirable data [2]. Extracts the features through compressive sensing and sparse representation to construct the model more fast and robust, and applied the LS-SVM classifier act as the optimization algorithm to separate positive templates from negative samples, finally hypergraph methodology use to improve the tracking accurately [3].

Majority of the object detection techniques are restricted to only a limited object group, the classification of datasets and of images will be a monotonous task. A training procedure called YOLO9000 was projected, This is a real-time object detector capable of detecting numerous items, including over 9000 samples [4]. In order to detects multiple objects, a Kalman filtering is secondhand, whose object structures will be arbitrary and Gaussian in fauna. Gaussian in personality. Once the location of the item transforms regularly and the movement is piercing, the Kalman filter is secondhand. Kalman filtering has two stages: 1. Prediction, in which the object location is predicted; and 2. Correction, in which the estimated value is approved in relation to the predetermined state assessment. The Particle Swarm Optimization (PSO) method is an extremely efficient object tracking technique that follows objects in moving video in a manner akin to bird flocking. The ideal object frame may be found based on the item's current location, its prior best position, and its best position in the overall frame.

Sparse representation object model is the majority and most commonly used in object recognition, two types of imaging sets were observed to apply. One is for object patterns, which have pixel values from previous frames. Another is inconsequential templates, which include noisy pixels. The pixels that make up object patterns are sparse manner described in a low-dimensional sub space, resulting in a path of non-zero entrances in the sparse medium, which represents the location of the pixel in a precise image frame [5]. However, the object model with thin depiction grieves from heavy obstruction and cannot account for things that come and disappear over time.

Object tracking in video clips is a problem for the available object tracking techniques. For applications ranging from surveillance to robotics and autonomous systems, the capacity to reliably detect and monitor objects in the midst of barriers, changing illumination conditions, object motion, shaking, and rotations is essential. Occlusions, scale shifts, and visual alterations are some of these difficulties. To accomplish this, need to overcome the above-mentioned issues, the paper has proposed CNN based FCT (Fast Compressive Tracking) method in which the sparse random matrix generated multiple object features is fed to the CNN model. The input is fed as two successive image frames, as the predominant features are extracted through the processing of stack of CNN layers and fed to the output layer. In order to improve object representation accuracy, compressive sensing techniques are included. The purpose of the Kalman Filter is to reduce uncertainty, remove erroneous data, and provide adaptation to irregular observations. At the output of the CNN model, Kalman filtering and Particle Swarm Optimization (PSO) is applied in order to locate the locations of the target image features. Particle Swarm Optimization is included to maximize forecasting accuracy and increase tracking efficiency. When these methods are combined, a strong solution that can handle problems like intricate movement patterns, occlusions, and visual imbalances is produced. A Support Vector Machine (SVM) classifier is trained in order to classify the target features of the images from the rest of the images. The proposed algorithm was able to track and detect objects under severe occlusion as well as was able to track images which are inconsistent in nature. CNN used extensively in the area of visual object tracking, with a strong ability to learn distinctive image features depictions, feature map selection methods to select discriminative features and discard noisy or unrelated ones. A multi-domain learning framework, called as MDNet, full network trained offline, and the connected layers counting Online fine-tuning of a single domain-specific layer [26]. CNN focused on visual recognition tasks, contains domain adaptation, fine-grained based recognition, and largescale scene recognition. The potentiality of a visual recognition system to attain elevated classification accuracy on exercise with sparse labeled data has shown to be a long term objective in computer vision research [6].

More reliable and effective multi-object identification techniques are desperately needed in dynamic and complicated situations, which is what inspired the proposed work. Situations where things are obscured, arrive suddenly, or vanish are difficult for traditional object detection techniques to handle. The CNN-FCT Methodology is selected as a ground-breaking solution in response to these constraints because of its exceptional characteristics and capabilities. To solve the issues with previous approaches, the CNN-FCT Methodology combines Convolutional Neural Networks (CNN), Particle Swarm Optimization (PSO), and Kalman Filtering. The model can recognize intricate patterns and representations in a variety of circumstances thanks to the efficient feature extraction provided by CNN. This is especially important to overcome occlusion-related difficulties, when standard approaches might not be able to recognize objects with enough accuracy. Dynamic optimization and tracking techniques are introduced via the combination of PSO and Kalman Filtering. PSO improves the model's ability to adjust to abrupt changes in the scene by improving the predicted positions of objects at each iteration. This is enhanced by Kalman Filtering, which offers a predictive filtering process that makes tracking more accurate particularly when objects appear and disappear. The shortcomings of conventional object detection techniques have been exposed by recent developments in computer vision and machine learning, particularly in situations that are extremely dynamic and unexpected. These issues are acknowledged in the suggested study, which creatively integrates CNN Compressive Sensing, PSO, and Kalman Filtering into the CNN-FCT Methodology. By doing this, the research hopes to greatly improve the performance of multi-object detection systems, strengthening their ability to withstand the intricacies of the real world and advancing computer vision applications across a range of industries.

The key contribution of the paper is given as follows:

- The proposed approach differentiates out since it integrates several innovative techniques. The system integrates CNNs, Compressive Sensing, Kalman Filter integration, and PSO to enhance object tracking effectiveness.
- An effective and precise item categorization and identification system is facilitated by the methodology's early integration of Convolutional Neural Networks (CNNs). This section makes sure that objects are located precisely inside of video frames, providing a solid basis for tracking operations.
- The accuracy of object representations is further improved by the application of compressive sensing methods. This addition improves object representation accuracy by effectively extracting relevant information from the video frames.
- By removing false data and lowering uncertainty in the tracking process, the inclusion of Kalman Filtering gives flexibility to erratic observations. By allowing the technique to adjust to erratic variations in object movement, this contribution enhances the methodology's durability and offers a more precise and dependable tracking mechanism.
- Particle Swarm Optimization is included to improve forecasting accuracy and tracking efficiency. This optimization process helps to maintain object tracking accuracy even in scenarios with occlusions, appearance modifications, and complicated motion patterns.

The remaining sections of the paper are ordered as follows: Section II deliberates about various multi-object tracking algorithms, Section III portrays about problem statement, Section IV discloses about the kernel filtering and sparse representation model, Section V gives a detailed explanation about the proposed CNN-FCT model, and Section VI depicts the investigational consequences and accomplishes the paper.

II. Related Works

As evidenced by this literature review, there are several visual object tracking methods available, we are going to assess few algorithms and methodologies, which have been extensively used for exploring reason.

A. Fast Visual Tracking

Li and Wang has proposed Dense Spatio-temporal Context Learning for fast graphic tracking [7]. This method creates a Bayesian framework, for associating the spatio-temporal aspects among the goal object and the foreground and circumstantial images pertaining to the image frames. For rapid learning and detection, this work applied four- Fast Fourier Transformation (FFT) and the results showed that this framework performed well against state-of-the-art techniques with admiration to effectiveness, correctness and robustness.

Bern et, al. [8] has proposed a incremental model algorithm, in which the target obejct is tracked increamently to a low-dimensional subspace presentation and also adapts to the dynamic changes pertaining to the target object. The increamenting algorithm is based on the principal component analysis and forgetting factor, that helps in enhancing the overall performance of the proposed tracking algorithm.

Najeeb and Ghani has made comparative study on the techniques of object tracking in various applications pertaining to computer vision domain namely, traffic surveillance, robot navigation and so on [9]. This paper studied about various object tracking methods, which tracks single or multiple objects in motion form a video sequence. The major object tracking techniques discussed in this survey are kernel tracking, point tracking and Silhouette tracking algorithms.

Dr. D. S. David [10] has proposed innovative and efficient object detection and tracking algorithm that utilizes optical flow in combination with motion vector assessment for object discovery and trailing in a series of frames. The optical flow exposes the details of the moving object and motion vector assessment provides the position of an object from successive frames and helps in increased accuracy in spite of motion blurs and cluttered image background.

B. Particle Swarm Optimization (PSO)

Moussa and Shoitan in this work has implemented Sequential Particle Swarm Optimization (PSO) for visual tracking, which tracks the objects by including significant features of the tracked objects [11]. This method resulted in increasing tracking performance because of its parameters that is flexible according to the fitness value of the objects and predicts the object's location correctly when it is in motion.

Nedjah has proposed an object tracking algorithm, which utilizes PSO technique, in order to detect the target object in motion from a video image series [12]. The entire video sequence is searched as an individual swarms which provides optimal solution.

C. Compressive Visual Tracking

Chen et al. [13] has proposed a scheme along with weight maps called as multi-sparse measurement matrix, which compresses the image but at the same time maintains the originality of the image features. The weight map merges a distinct weight as well as a characteristic weight to proficiently differentiate the aimed manifestation and position. Furthermore, a dispersion function is applied for the digital updating of the intended template letting to track together the position and extent of the target object.

Nguyen et al. [14] proposed a method which holds two steps one is effective Cholesky decomposition on GPU and FPGA. In order to improve the performance with respect the memory access, second is CS signal reconstruction applied on FPGAs and GPUs which helps to reduce the computation.

Li et al. [15] has proposed template matching dealing out phase along with the compressive tracking method, in demand to maintain the constant frame frequency and rise the strength of the tracker. This template comprises of all the available data and similarity metrics have been used to compare the available data with the regularly used images series.

D. Kalman Filtering

Kaderali in this work has implemented IMM cubature Kalman filter (IMM-CKF) in order to track the orbiting space objects [16]. This study utilizes the geometric association among the planetary object specifically space craft, space based optical (SBO) sensor, and the sun for tracking the space object. A situation which encompasses the aimed spacecraft and four SBO sensors is utilized to check efficiency of the IMM-CKF. A comparison was made between IMM-CKF and the normal cubature Kalman filter (CKF). The outcome designates that IMM-CKF is extremely vigorous than the CKF when the space object experiences a movement. Shantaiya et al. [5] has proposed multiple objects tracking from the video series using optical flow algorithm. This algorithm uses Kalman filtering to detect and track the moving objects in each frame, which helps in identifying moving objects with occlusion, blurred objects and so on. Ullah et al. [17] for the usage against non-linear systems proposed a novel method known as Unscented Extended Kalman Filtering (UEKF). This new method of Kalman filtering is same as traditional Kalman filtering except, for every running non-linear sample, the deterministic sample is caught with non-linear mean and linear covariance. This new method offered better performance when compared to that of conventional Kalman filtering.

E. Convolutional Neural Network-based Visual Tracking

Xiao and Pan has researched, how CNN filers helps in tracking moving objects by increasing the depth of the CNN [20]. It was found that, by making the depth to 16-19 layers, the proposed ConvNet showed better performance than the traditional visual tracking methodologies when evaluated against state-of-arts results. Mahmoudi et al. [18] has proposed CNN based online moving object tracking algorithm, in which the tracking missed error rates are reduced by employing truncated structure loss function and temporal selection mechanism for discriminating positive and negative samples. Liu et al. [19] has proposed online CNN model for tracking visual objects. For hierarchical feature learning, and to reduce the dependability of CNN, k-Means is applied. Regression model is involved to notice the positive samples of the goal visual object. The investigational outcomes showed that, the planned model executed very well when compared to another Online CNN model. Li and Yang proposed a model based on rich feature hierarchies of CNNs cultured from a large-scale datasets and train a liner correlation filter on each convolutional layer and conclude the goal location with a coarse-to-fine searching method [20], yields outcomes favorably against the state-of-the-art approaches in relations of robustness and accuracy.

Object tracking techniques in computer vision, Incremental Model Algorithms, Dense Spatio-temporal Context Learning, PSO for visual tracking, Compressive Visual Tracking with weight maps and efficient signal reconstruction, Kalman Filtering for tracking orbiting space objects, and CNN-based Visual Tracking are just a few of the advanced techniques for visual tracking that are examined in this literature review. To improve tracking performance, these approaches use a variety of strategies, including CNN filters, optical flow, Bayesian frameworks, Principal Component Analysis, and Cholesky decomposition. The review illustrates their uses in various contexts and shows how well they work to increase accuracy and resilience when compared to state-of-the-art outcomes and conventional tracking approaches.

Numerous strategies have been investigated by researchers, such as incremental model algorithms, Bayesian frameworks, comparative examinations of tracking methods, novel object identification and tracking algorithms, and the use of optimization strategies like Particle Swarm Optimization (PSO). Also, there is an emphasis on improving tracking performance by utilizing deep learning techniques, namely CNNs, multi-sparse measurement matrices, compressive tracking, and Cholesky decomposition. The main difficulty, despite the variety of approaches, is creating dependable and effective object tracking systems that can deal with things like motion blur, crowded backgrounds, and shifting object appearance.

III. PROBLEM STATEMENT

From the above discussed literatures, it is found that the challenge involves tackling the difficulty of multi-object tracking in dynamic environments, where a large number of objects must be reliably and precisely tracked despite occlusions, visual discrepancies, and intricate movement patterns [21]. The goal is to improve tracking accuracy and resilience by combining CNNs, PSO, Kalman Filtering, and Compressive Sensing. The issue statement's primary objective is to develop and evaluate an improved multi-object tracking system that employs these techniques in order to progress object tracking and computer vision applications and achieve high tracking accuracy and reliability in challenging circumstances. The method used in the study is called Multiple Object Tracking, or MOT. In order to overcome the aforementioned problems, the current study establishes a dependable and real-time tracking device that is capable of handling the difficulties associated with multiple object tracking in a variety of instances. This is achieved by using an effective compressible tracking infrastructure based on CNN models in conjunction with Kalman filtering and PSO algorithm.

IV. PROPOSED CNN-FCT MODEL FOR OBJECT TRACKING

The CNN-FCT Methodology combines a number of essential elements, each of which has a unique function in tackling the difficulties involved in multi-object recognition in dynamic and complex situations. The effectiveness of the methodology is largely due to its novel features, which include the use of Particle Swarm Optimization (PSO) to improve predictions, Fast Compressive Tracking (FCT) for object tracking, Segmentation with Otsu thresholding, Kalman Filtering, and Convolutional Neural Networks (CNN) for object detection and classification. This all-encompassing strategy outperforms separate approaches and has promise for robotics, autonomous systems, and surveillance. Fig. 1 depicts the proposed CNN-FCT methodology.

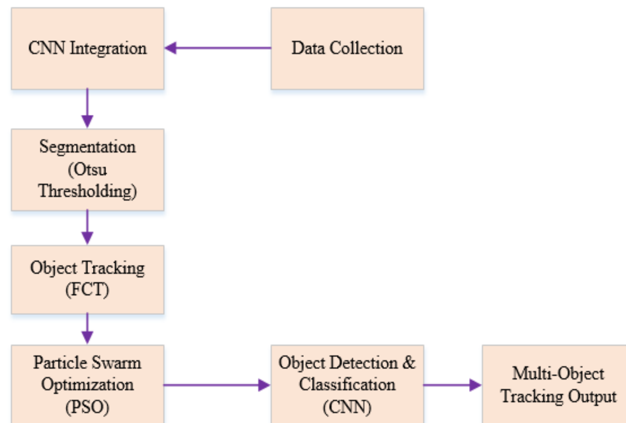


Fig. 1. Proposed CNN-FCT Methodology.

A. Data Collection

The MOT challenge 2017 dataset is used for example video clips in order to evaluate the suggested approach. In all, 11245 frames are included in its 14 series of detections, seven of which are used for testing and the other seven for training [22].

B. Kalman Filtering

In the predictive tracking and filtering process, Kalman Filtering is an essential part. Kalman filtering improves the estimated item positions using dynamic measurements and predictions by implementing a recursive method. This guarantees, even in situations where objects appear, disappear, or are obscured, that their trajectories are monitored precisely. The Kalman Filter is used to improve object tracking accuracy. The Kalman Filter gives an ideal estimate of the object's state, efficiently minimizing the effects of noise and uncertainties, by integrating noisy and uncertain observations with predictive models. As fresh measurements come in, it continually updates its estimation in order to dynamically respond to changing conditions. Smooth and precise tracking is made possible by the Kalman Filter's capacity to combine historical data and forecasts with current measurements, especially in conditions characterized by noise, occlusions, and fluctuations in object speed. It is essential to attaining accurate and consistent object tracking results because of its iterative process of prediction, update, and state estimation, which acts as a strong mechanism to preserve the trajectory and location of the tracked item. Kalman filtering, which is especially effective at managing noisy data, helps to lower the number of false positives and negatives and ensure accurate estimations of the object state. The system is robust in dynamic circumstances because of its capacity to adapt to changing surroundings and retain continuity between frames. This is made possible by its sequential frame processing capabilities. By offering temporal context for object tracking when combined with CNN Compressive Sensing, Kalman Filtering enhances the capabilities of CNNs and produces a more precise, accurate, and real-time multi-object detection system that performs well in demanding and dynamic settings.

C. Segmentation

To successfully separate items from the background, the segmentation module uses Otsu thresholding. By separating foreground and background pixels with the best possible threshold, this method enhances object localization. Otsu thresholding improves the segmentation process, making it possible to extract features for later stages of the approach with greater accuracy. Otsu thresholding, a popular image segmentation approach, distinguishes between object areas and background in a video stream, which is crucial for object tracking. With the help of automation, the ideal threshold value that reduces intra-class variation in object and background areas may be found. By transforming grayscale or color frames into binary images, where pixels are classed as either object or background based on their intensity values, Otsu thresholding efficiently separates items of interest in object tracking. Otsu thresholding allows subsequent object tracking algorithms to precisely detect and follow the movements of objects over time, improving the accuracy and robustness of the tracking process. This is accomplished by segmenting the video frames into discrete zones. It is given in Eq. (1) below.

$$\sigma_z^2 = W_x\sigma_x^2 + W_y\sigma_y^2 \tag{1}$$

In this equation, W_y stands for the weight of the frontal image and σ_y^2 for its variation, whereas W_x stands for the weight of the background image and σ_x^2 for its variation. With the use of this approach, researchers may determine which pixels in an image serve to balance the foreground and background, as well as how many of them overall there are in comparison to both the background and foreground pixels. The average and variance are then calculated for the suitable backdrop and foreground. The weight and variation were then used to establish the various thresholding level values.

D. FCT for Object Tracking

For effective and real-time object tracking, Fast Compressive Tracking (FCT) integration is essential. Using compressive sensing techniques, FCT is able to track and rebuild objects over several frames. This method offers strong tracking capabilities and is especially useful in situations where objects are obscured or move suddenly. Feature correspondence and transformation techniques are used in FCT, a unique approach to object tracking, to improve the accuracy and resilience of tracking in complicated visual situations. In order to build object correspondences, this technique first extracts and matches characteristic features from successive frames. An adaptive transformation procedure then adjusts for variations in object appearance, size, and orientation. FCT is a potential approach for multi-object tracking in real-world applications because it combines feature-based matching with transformation-based modelling to address issues including occlusions, illumination changes, and object deformations. By offering a more robust and flexible framework for monitoring objects across a variety of applications, including robotics, autonomous systems, and surveillance, this method has the potential to substantially enhance the fields of computer vision and object tracking.

Fast Compressive Tracking (FCT), a ground-breaking technique for reliable and effective object tracking in video sequences, has gained prominence. FCT presents a revolutionary method for collecting and describing the look of objects, building on the concepts of compressive sensing, enabling real-time tracking even in difficult circumstances. FCT makes use of a condensed image of the object's appearance. Online learning of this condensed model is effective and adapts to changes in object appearance and background noise. The distinguishing strength of FCT is its robustness to occlusions, changes in illumination, and deformations while still successfully separating the object from the backdrop. FCT is particularly suited for high-speed applications like robotics, surveillance, and interactive systems since it can achieve amazing tracking speeds while drastically decreasing computing overhead. Compressive sensing's effective signal capture combined with FCT's adaptable appearance modelling results in a significant improvement in object tracking. Algorithm 1 explains the object tracking.

Algorithm 1: Multi-Object FCT Algorithm

Input: A representative model from Y^t, Y, t_{th} frame.

Step 1: retrieve $v(Y^t)$ by selecting the searching range " $\kappa_y \geq 1$ and looking for step ꚍs≥ 1".

Step 2: classifier CNN equations are used, the tracking location $I_{(t-1)}$ is achieved with the best predicted outcome.

Step 3: retrieve $v'(Y^t)$ by selecting the searching range " $\kappa'_y \geq 1$and $\leq \kappa_y$ and looking for step ꚍ's≥ 1 and≤ ꚍs".

Step 4: By using $v'(Y^t)$ to CNN equations, the tracking location $I_{(t)}$ is achieved with the best predicted outcome.

Step 5: Now, extract the features from the two sets $v\,(Y^t)$ and $v'(Y^t)$.

Step 6: Apply the sampled feature to the SVM classifier in the SVM equation now.

Output: Object is tracked at $I_{(t)}$.

E. Object Detection and Classification with CNN

Utilizing Convolutional Neural Networks (CNN), the approach performs exceptionally well in object classification and feature extraction. Because CNNs are skilled at extracting hierarchical representations from incoming data, the model can recognize complex patterns and features. This improves the methodology's ability to recognize and classify objects accurately, making it appropriate for a wide range of item kinds and intricate situations. The way of recognizing and classifying things inside images and video frames has been completely transformed by the state-of-the-art computer vision. CNNs are able to identify things with great precision because of their inherent ability to automatically learn hierarchical properties from input. Using this method, a CNN model is trained on large datasets to identify particular classes or objects. The trained CNN is used to input images during the inference phase, when it searches the whole scene for objects and then gives them class labels. Numerous real-world uses for this technol8ogy exist, such as in driverless cars, spying, the interpretation of medical images, and more. Its capacity to manage intricate situations like object occlusions and changing object sizes and orientations makes it a potent instrument in the field of computer vision, opening the door for sophisticated applications in a variety of sectors.

CNNs are used for object identification and classification in computer vision applications, such as multi-object tracking and real-time situations. These applications benefit from hierarchical feature extraction, spatial hierarchies, and robustness to fluctuations. CNNs are an invaluable tool for applications requiring precise and effective object detection in a variety of visual settings because to their versatility and adaptability. In the field of object tracking, CNNs are used to locate and identify objects inside video frames, giving not only their locations but also the labels that go along with them. Even in a variety of tough visual environments, the CNN can reliably distinguish between multiple object categories thanks to its capacity to learn intricate features and patterns from large training data. CNNs are adept at classifying objects; they give labels according to traits they have learnt and are adaptable to different categories. Training is accelerated by utilizing transfer learning, particularly in situations when task-specific data is scarce. With improved designs like SSD or YOLO, CNNs find applications in real-time activities and exhibit robustness to fluctuations in object appearance.

The input is fed to the CNN network, as a combination of two cropped regions from two next frames as explained in Algorithm-2. Then it is propagated through five layers of CNN layer, where researchers remove all the remaining layers from the pooling layers to store only precise information of the cropped pair of input images. Researchers produce a couple of frames (CF_t, CF_{t-1}) with target center $\left(a+\frac{w}{2}\,;b+\frac{h}{2}\right)_{t-1}$,and size $(p_w\,,p_{sh})_{t-1}$, where, $t-1$ stands for the directory of preceding frame and p describes the exploration range.

The collected areas are resized into $m_i * m_i$ with scale factor $(sx;sy)$ which is well-defined beforehand affording with the real size of images, and nourished into a CNN model to attain the 1^{th} Convolutional beginnings. Tracking information linked list is presented to discover novel substances of attention incoming the image which is a sequence of protuberances N depicting noticed objects which emerges in frames with the same ID from the initial to the final. Each node N_t comprises the objects in CF_t. Every object has ID which is identified by its aspects. When the similar object emerges over again, it goes on the olden ID in its place of a new one. Hence, the tracking information linked list can be applied to resolve obstruction issues and to forecast misrepresented objects. The suggestion is made by the PSO algorithm.

Algorithm 2: CNN based Classification Algorithm

Input: A pair of cropped frames $(CF_t;CF_{t-1})$ target center $(a+\frac{w}{2};b+\frac{h}{2})_{t-1}$, and size $(p_w;p_{sh})$

Step 1: Resize the cropped pairs $m_i * m_i$ with scale factor $(s_x;s_y)$

Step 2: The resized cropped image is fed to the CNN network to obtain 1^{th} convolution activation

Step 3: The precise features are extracted from the cropped inputs by removing all the images in the pooling.

Output: Classified object linked list obtained with Node N_t and associated objects O_t with ID for every node.

F. PSO for Enhancing Predictions

To improve and refine the predictions produced by the approach, Particle Swarm Optimization (PSO) is utilized. Accuracy and flexibility are enhanced through parameter optimization of the model by PSO. Ensuring that the model dynamically adapts to changing conditions is especially useful in situations when abrupt changes take place. PSO is essential for object tracking since it improves predictions after CNN classification. Following the CNN's classification and identification of the objects in the video frames, PSO enters to improve the original predictions by maximizing the predicted object locations and trajectories. By utilizing PSO's optimization skills, the tracking algorithm may fine-tune the anticipated item positions based on elements like motion models, historical data, and real-time measurements. This improves the initial forecasts' accuracy and allows them to be adjusted to the tracking scenario's dynamic nature. In order to account for uncertainties and imperfections caused during object categorization and initial tracking, PSO's repeated optimization procedure refines the anticipated locations. Its combination with other methods, such CNN-based object recognition and Kalman Filtering, results in a hybrid strategy that provides a thorough plan for precise and adaptable forecasts in multi-object tracking situations. Real-time applications are made easier by PSO's computational efficiency, which guarantees fast and rapid prediction generation. This is important in dynamic contexts where timely and accurate forecasts are necessary for making wise decisions.

PSO is attracting more academics because of its versatility and resilience, particularly for challenges in dynamic environments, it is a population-based stochastic optimization technique. Idea of PSO is the particles fly around randomly over the image searching for the best value in each round. The fittest value of the particle in its entire search at every round is called P_{Best}. The fitness function is applied to all particles, and the fitness value (best solution) is estimated and stored. The fitness value of the current optimal particle is referred to as "P_{best}". PSO maximises the best population value gained as far as any particle in the neighbourhood, and its location is known as l_{best}.

When all of the generated populations are measured as topological neighbors by a sensitive particle, the best value is chosen among the executed population, and that sensitive best value is recognized as the best solution as g_{best}. The PSO is always attempting to change the speed of each particle towards its p_{best} and l_{best}. The speed is determined by arbitrary terms, which have arbitrarily generated quantities for the speed towards the p_{best} and l_{best} locales. The preceding stages are depicted as an Algorithm 3.

Algorithm 3: PSO Algorithm for fast visual tracking objects

Input: at t_{th} frame of the image.

Swarm size (N), Maximum number of Iteration (M), L_rate(L_r) , acceleration coefficients (C_1) ,(C_2), prior width (w) and height(h)

Initialize the image size (n)

For $i = 1\ to\ n$ // $i = 1, 2, 3 \dots \dots n$// starting with 1

Initialize the swarm particles randomly $SP_j = [\, X_{s1}, X_{s2}, X_{s3}, X_{s4} \dots X_{sn}]$

// $j = 1,2,3, \dots . N_{sp}$ $j = 1,2,3, \dots . N$, $X_1 = [x, y, w, h]$, x– starting x-axis, y-starting y-axis

Calculate the fitness of each particle (i.e., compute the ratio classifier for all the samples

determine the p_{best} and g_{best} for the current round (“iteration”).

While $(t < M)$

for $j = 1{:}\, N_p$ //starting with_2

evaluate and update the velocity of the particles

$$V_j(t+1) = w(t+1) * V_j(t) + C_1 * rand * \left(P_{best} - SP_j\right) + C_2 * rand * (g_{best} - SP_j)$$

$P_j(t+1) = p_j(t) + velocity\ (t+1) * lrate$ // “revise updated position”

end for // “end for 2”

Calculate the fitness of particles (new_fit)

revise the P_{best} and g_{best}

// P_{best}

if $(new_{fit} > P_{best}\ fit)$then

$P_{best}\_sol = P_j$;

$P_{best}\_fit\ = new\_fit$

else

do nothing

end if

// G_{best}

if $g_{best}\_fit < max(new\_fit)$) then

$g_{best}\_Sol = P_j(index\ of\ Max(new\_fit))$

$g_{best}\ Fit = Max(new\_fit)$

end if

end while // Repeat until the maximum num. of iterations is reached.

End for // until n frames are reached.

V. Experimental Results

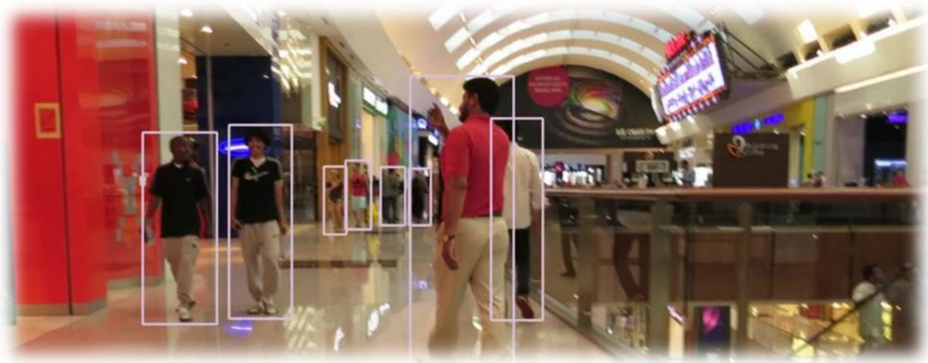
To assess the proposed methodology, MOTchallenge 2017 dataset is utilized for sample video clippings. It contains encompasses 14 series of detections, seven for training and seven for testing, thereby on a whole comprising 11245 frames. The investigations are executed in OpenCV environment, in which python is applied as the programming language.

A. Tracking Samples of the Proposed CNN-based Fast Compressive Tracking Methodology

1) Detection of people in a busy shopping mall- sample video for complete occlusion: The detection people in a shopping mall in highly occluded condition are depicted in Fig. 2(a) – (d). The video sample consists of 30 Frames Per Second (FPS), with 75 tracks and 12389 detection boxes. From Fig. 2(a) and (b) researchers can show that, the proposed methodology was able to track and detect people whose motion is unpredictable.

From Fig. 2(c) and (d) it can discovered that, the man in blue shirt has completely occluded a small boy whose shoes are only visible in the Fig. 2(c). In the frame i.e., Fig. 2(d), the boy is easily detected by the proposed methodology from which researchers can give assurance that the novel methodology is capable of detecting objects even in complete occlusion.

2) Tracking people in busy street- sample video for sudden arrival and fading of the object: Fig. 3(a), 3(b) and 3(c) depicts the example of sudden arrival and fading of the goal object. This sample video comprises of 30 FPS, with 83 tracks and 47557 detection boxes.



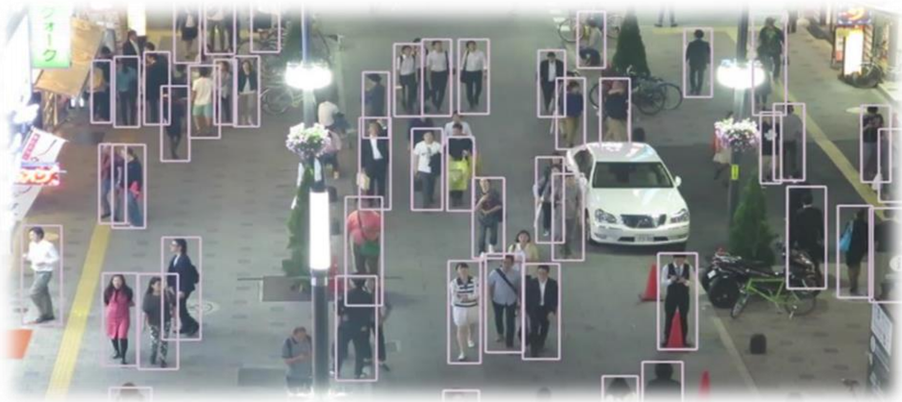
(a)

(b)

(c)

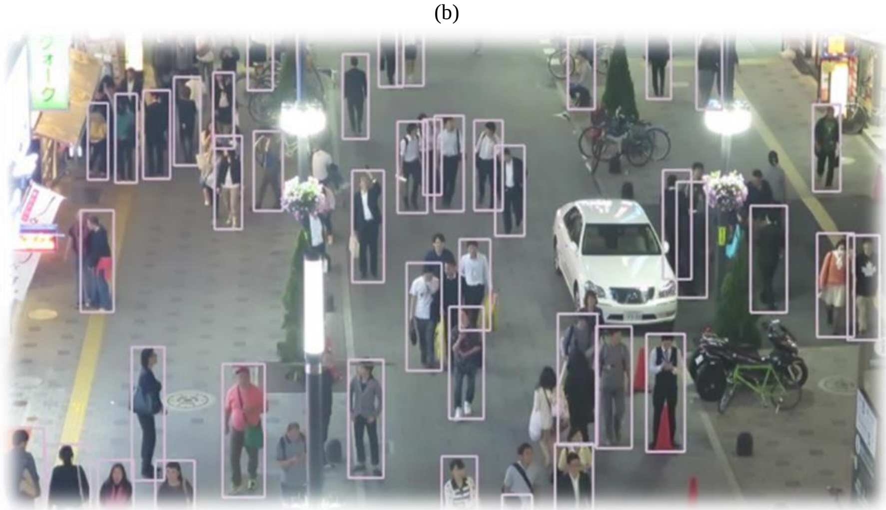
(d)

Fig. 2. (a) – (d) Detection of people in shopping malls at various frames.



(a)

(b)



(c)

Fig. 3. (a)-(c) Example of sudden appearance and disappearance of target object.

From the Fig. 3(a), it can be found that in the third street lamp, a man wearing a black hat is standing near the post. In the next Fig. 3(b), he was hidden behind the post. After sometimes, he is again appearing and tracked in the Fig. 3(c). Thus the proposed methodology was able to track objects with sudden appearance and re-appearance behavior.

Researchers are evaluating using MOT as a performance metric suite, which comprises MOT_A (Multi-Object Tracking Accuracy) and MOT_P (Multi-Object Tracking Precision) as the chief metrics The MOT_A collects "three error causes: False Positives (FP), False Negatives (FN) and Identity Switches (IDs)". "*Gt* is the sum of the goals in each edge. Researchers also investigate about the sum of courses of Ground_Truth (GT), Mostly_Tracked (MT), Mostly_Lost (ML), Fragment (FM), Partially_Tracked trajectories (PT), and Multiple Object Tracking Accuracy with log 10 (IDs) (MOTAL), False_Alarms per Frame (FAF). Researchers also acknowledge the assessment metrics Recall (Rcll) and Precision (Prcn)". Each technique's rate is restrained in frames per second (FPS). The subsequent table compares several object followers in the MOT2017 encounter dataset to the planned scheme.

The above Table I make a comparison between the existing multi-objects tracking to that of the proposed methodology. The comparison was made taking two datasets which was used as sample video sequences for the experimental purpose namely, shopping mall dataset and busy street dataset. From the table it can be concluded that, the proposed methodology showed better performance against MOT2017 dataset state-of-arts methods when compared to that of existing object tracking methodologies.

The graphical depiction of the comparison between the proposed CNN-FCT and M-FCT is shown in the following graphs. Fig. 4 compares the Accuracy, Robustness and efficiency of the proposed methodology with that of M-FCT. It was found that, CNN-FCT showed enhanced performance measures when compared to that of M-FCT.

Fig. 5 depicts the MOT accuracy between CNN-FCT and M-FCT, whereas Fig. 6 depicts the Most Tracked and Most Least measures of the proposed and existing methodologies. As the proposed CNN-FCT tracks object even if the object is fully occluded or appears and re-appears randomly, the MOT Accuracy is higher than that of M-FCT. Even though M-FCT is better when compared to conventional object tracking methods, its track loss rate is higher when compared to that of CNN-FCT. Hence from the comparisons, it can be evident that the proposed methodology performed well against all recently proposed multi-object tracking algorithms, especially in case of complete occlusion and object appearance and disappearance scenario.

B. *Performance Metrics Evaluation*

Quantitative indicators that evaluate accuracy and effectiveness are referred to as metrics of performance. These measurements are employed to assess and contrast various tracking approaches or algorithms. Here are a few performance measures that are frequently employed in object tracking such as recall, mAP, F1-score, Peak Signal to Noise Ratio (PSNR) and Root Mean Square Error (RMSE).

1) Recall: The proportion of real optimistic outcomes that a model correctly organizes as being optimistic is known as recall. The proportion of true positives to the entire of true positives and false negatives is secondhand to calculate it. It is given in Eq. (2) below:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \qquad (2)$$

TABLE. I MOT2017 CHALLENGE DATASET MULTI-OBJECT TRACKING VERSUS THE PROPOSED METHODOLOGY

Datasets	Methods	MOT-A	MOT_P	MTrack	MLost	FP(+ve)	FN (-ve)	IDs (/Rcll)	FM(/Rcll)	Speed
Mall dataset	ACF [30]	33.7	76.5	7.2%	54.2%	5,804	112,587	2,418 (63.2)	2,252 (58.9)	1.3
	ZF [30]	33.2	75.5	7.8%	54.4%	6,837	114,322	642 (17.2)	731 (19.6)	0.3
	CNN-FCT (proposed)	**40.9**	**88.8**	**9.8%**	**34.0%**	**3,356**	**83,108**	**567 (11.78)**	**670 (18.4)**	**12.5**
Busy Street dataset	ACF [30]	29.7	75.2	5.3%	47.7%	17,426	107,552	3,108 (75.8)	4,483 (109.3)	0.2
	ZF [30]	26.2	76.3	4.1%	67.5%	3,689	130,54	365 (12.9)	638 (22.5)	22.2
	CNN-FCT (proposed)	**38.4**	**90.34**	**9.7%**	**36.0%**	**3,378**	**85,108**	**572 (12.78)**	**620(18.4)**	**13.6**

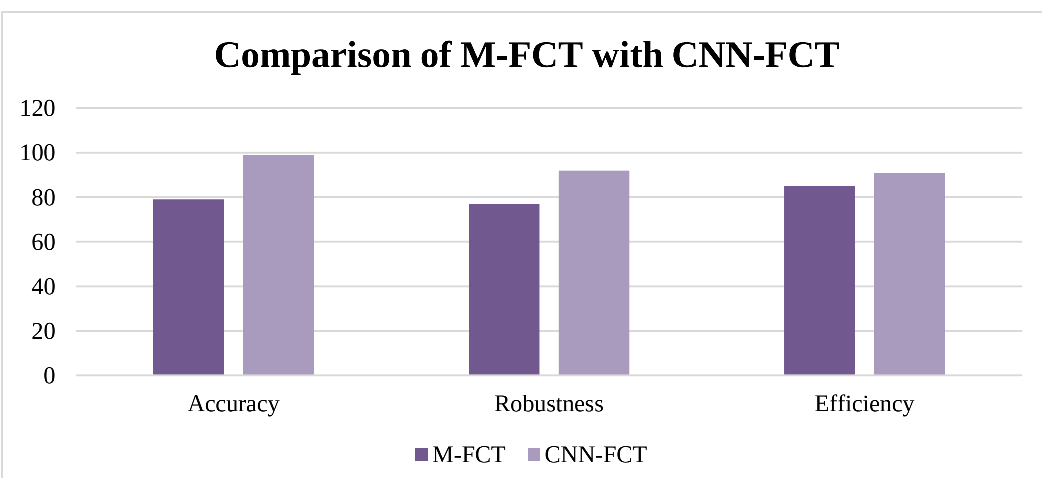


Fig. 4. Performance metrics comparison graph.

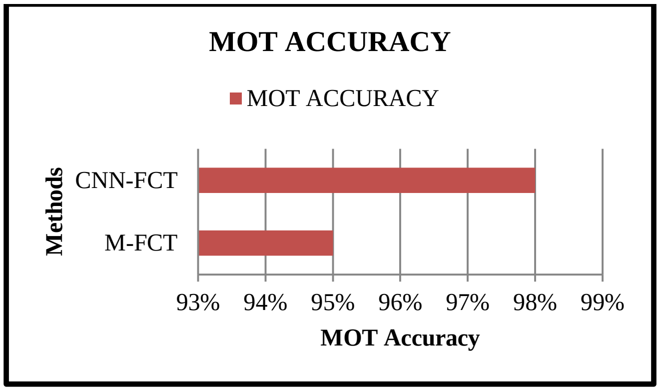


Fig. 5. Mot accuracy.

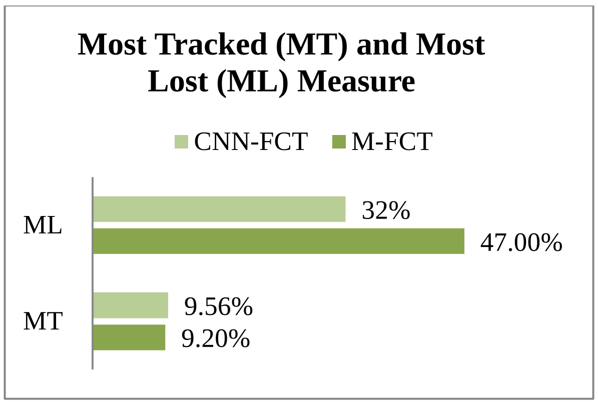


Fig. 6. ML and MT measures.

2) mAP: The mean over every category is then calculated by mAP after calculating the standard deviation of precision for every class. Its equation is given in Eq. (3) below,

$$mAP = \sum_{p=1}^{P} \frac{Average\ Q(p)}{P} \quad (3)$$

3) F1 Score: The F1 score is an individual statistic that syndicates precision and recall. It is the vocal average of these two metrics. It is frequently employed in binary categorization jobs where the proportion of both positive and negative instances is unbalanced. Its equation is given in Eq. (4) below,

$$\text{F1-Score} = \frac{2 \text{ x Precision x Recall}}{\text{Precision+Recall}} \quad (4)$$

4) Root Mean Square Error (RMSE): A standard metric for assessing the effectiveness of models of regression is RMSE. By taking into account the squared variations, it calculates the average variance among the expected and actual outcomes. When greater errors are more important, RMSE is especially helpful since it draws attention to more significant discrepancies. It is given in Eq. (5) below,

$$RMSE = \sqrt{\sum_{j=1}^{M} \frac{\|x(j)-\hat{x}(j)\|}{N}} \quad (5)$$

Here, j is represented as the variable; M is represented as the non-missing data points; $x(j)$ is represented as the actual observation time series; $\hat{x}(j)$ is represented as the estimated time series.

5) Peak Signal to Noise Ratio (PSNR): A commonly used statistic for assessing the effectiveness of video or image compression methods is PSNR. It calculates the difference among the highest possible signal strength and the background noise power. It is given in Eq. (6) below,

$$PSNR = \frac{10\ log_{10}\ (peak\ value)^2}{MSE} \quad (6)$$

TABLE. II COMPARISON TABLE OF RECALL, MAP, F1 SCORE

Method	Recall	Map	F1 score
Clustering Method [23]	71	53	61
YOLOV4 DNN Model [24]	93	96	93
YOLOV3 Model [25]	41	46	53
Proposed CNN-FCT Model	95	96.9	95.3

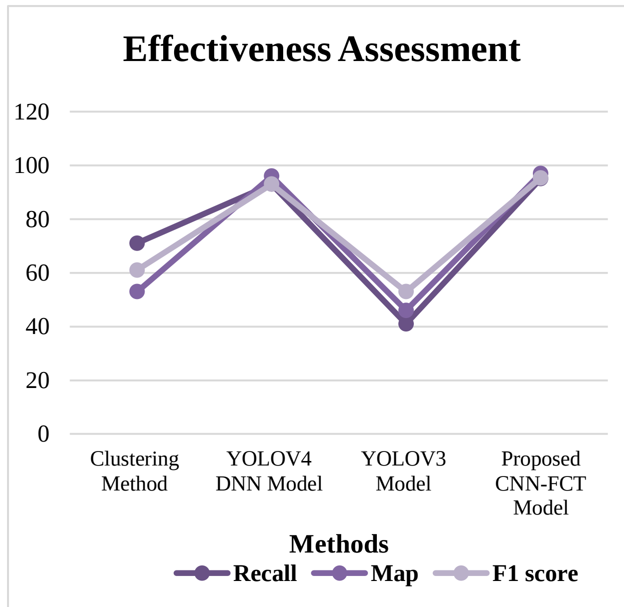


Fig. 7. Effectiveness assessment of Recall, mAP and F1-Score.

Table II makes the assessment among the existing methods recall, mAP and F1 score with the proposed CNN-FCT method. The proposed model produces greater recall, mAP and F1 score points and Fig. 7 depicts the comparison graph of the existing methods recall, mAP and F1 score with the proposed CNN-FCT method.

Other metrics like RMSE and PSNR also takes place and the below Table III shows the assessment table of it and Fig. 8 shows the comparison graph of the existing methods RMSE and PSNR with the proposed CNN-FCT method.

TABLE. III COMPARISON TABLE OF RMSE AND PSNR

Method	RMSE	PSNR
CNN Method [26]	24.09	30.38
OTSU Algorithm [27]	43.18	29.16
Proposed CNN-FCT Model	18.25	35.72

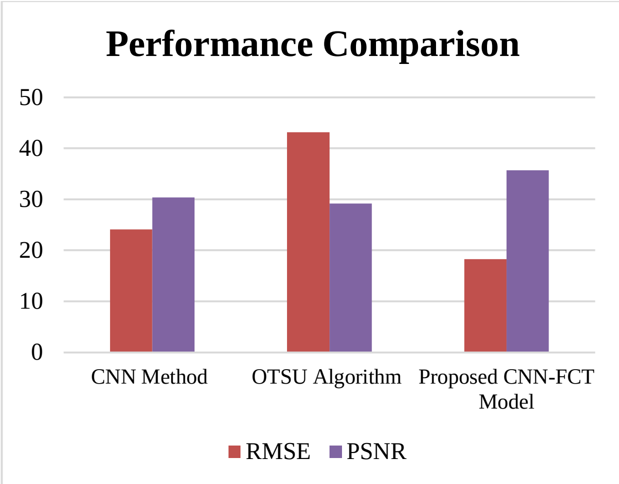


Fig. 8. Comparison graph of RMSE and PSNR.

The phrase fitness improvement over iteration in this context refers to the detection method's use of an optimization iteration procedure that enhances the precision and effectiveness of various item monitoring. A Convolutional Neural Network (CNN) model, which serves as an extractor of features for representing the objects in the video frames, is at the core of the approach. The CNN algorithm extracts the object's features after estimating the object's location in the initial frame during the tracking procedure. The predicted location of the first frame might not be completely precise, leading to a less-than-ideal initial solution. The Kalman Filter and PSO algorithm are used to modify the object's projected position at every iteration, and the fitness metric is calculated as well. This fitness statistic measures how closely the predicted position of the object matches the actual or intended position. Fig. 9 shows the fitness improvement graph.

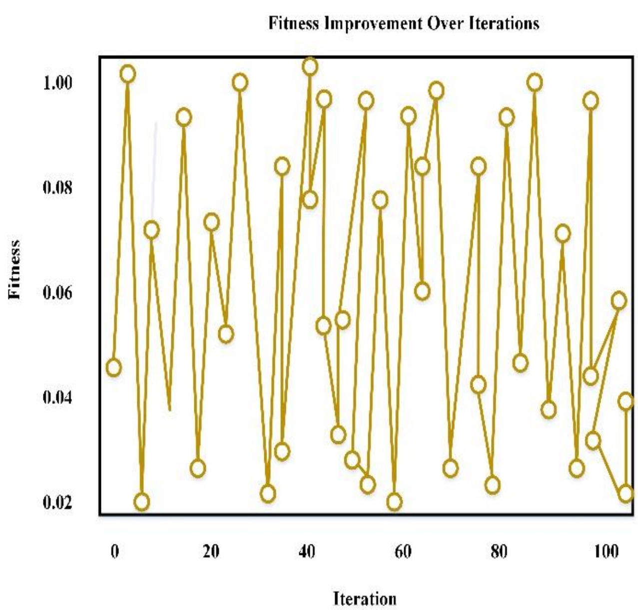


Fig. 9. Fitness improvement over iteration graph.

The graph contrasting the efficacy of a model with and without Accuracy, Recall, mAP (mean Average Precision), and F1 score optimization offers insightful information about the importance of focused improvement in particular machine learning tasks. The algorithm undergoes training in the with optimization case primarily deals with the goal of enhancing The key performance indicators for applications like object identification and classification are accuracy, recall, mAP, and F1 score. The graph shows a stronger upward trend as a result of the model's constant fine-tuning to maximize these metrics. When there is no optimization, the algorithm is trained without giving these metrics any special consideration, which results in more or less noticeable improvements in Accuracy, Recall, mAP, and F1 score. The graphs of accuracy, recall, mAP, and F1 score with and without optimization are displayed in Fig. 10.

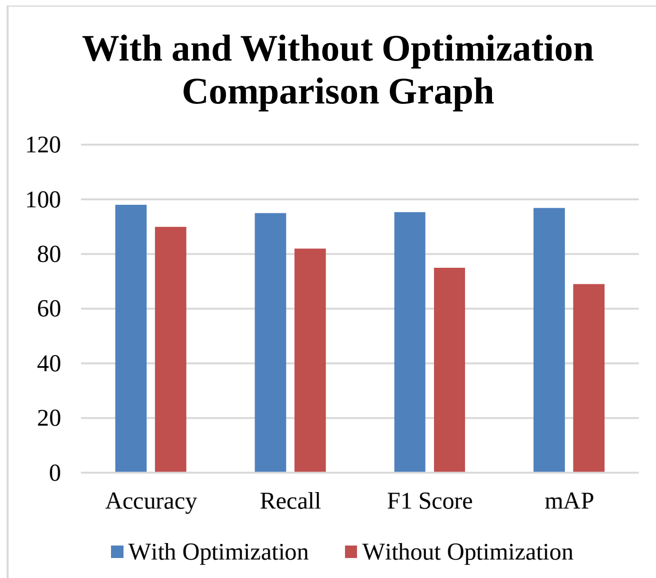


Fig. 10. With and without optimization comparison graph of accuracy, Recall, mAP, F1 score.

C. Discussion

In order to improve object tracking, the research article suggests a Fast Compressive Tracking approach based on Convolutional Neural Network (CNN) models. The CNN model is fed two sets of the dimensionally decreased object characteristics using this method. The CNN algorithm analyses the input frames via consecutive layers of the CNN model, removing the salient characteristics from the input frames. The Particle Swarm Optimization (PSO) technique is used by the CNN model's output coating to record the positions of the monitored targeted image features. The intended image is then classified using a Support Vector Machine (SVM) classifier based on the monitored positions. Here, the planned CNN based fast visual tracking architecture is depicted in Fig. 1. The Multi Object FCT algorithm is given in Algorithm 1. CNN based Tracking algorithm is given in Algorithm 2. Then, PSO algorithm for fast visual tracking object is given in Algorithm 3. In the results section, the detection people in a shopping mall in highly occluded condition are depicted in Fig. 2(a) – (d). The video sample consists of 30 Frames Per Second (FPS), with 75 tracks and 12389 detection boxes. From Fig. 2(a) and (b) researchers can show that, the proposed methodology was able to track and detect people whose motion is unpredictable. Then, Fig. 2(c) and (d) it can discover that, the man in blue shirt has completely occluded a small boy whose shoes are only visible in the fig. 2(c). Fig. 2(d), shows the boy is easily detected by the proposed methodology from which researchers can give assurance that the novel methodology is capable of detecting objects even in complete occlusion. Fig. 3(a), 3(b) and 3(c) depicts the example of sudden appearance and disappearance of the target object. This sample video comprises of 30 FPS, with 83 tracks and 47557 detection boxes. After that, Table I makes a comparison between the existing multi-objects tracking to that of the proposed methodology. The comparison was made taking 2 datasets which was used as sample video sequences for the experimental purpose namely, shopping mall dataset and busy street dataset. Fig. 5 depicts the MOT accuracy between CNN-FCT and M-FCT, whereas Fig. 6 depicts the Most Tracked and Most Least measures of the proposed and existing methodologies. Table II makes the comparison between the existing methods recall, mAP and F1 score with the proposed CNN-FCT method. The proposed model produces greater recall, mAP and F1 score points and Fig. 7 depicts the comparison graph of the existing methods recall, mAP and F1 score with the proposed CNN-FCT method. Table III displays the assessment table of it and Fig. 8 displays the assessment graph of the existing methods RMSE and PSNR with the planned CNN-FCT method [26] [27]. Fig. 9 shows the fitness improvement graph and it is used to determine a performance achievement's advantages or flaws. Fig. 10 shows the with and without optimization graph of Accuracy, Recall, mAP and F1 score because it is important to remember that these conclusions are hypothetical considering the limited facts at hand [23] [24] [25].

Compared to previous multi-object identification techniques, the CNN-FCT Methodology works well because it addresses these drawbacks. Real-time tracking may be difficult for traditional methods, particularly in situations where there is occlusion or abrupt object movements. In order to overcome these drawbacks, the CNN-FCT Methodology incorporates Kalman Filtering, FCT, CNN, and PSO. This results in improved accuracy, flexibility, and resilience while dealing with changing environmental circumstances. As a promising development in the realm of multi-object identification, the methodology stands out for its capacity to tackle a variety of obstacles.

The proposed CNN-FCT technology in addition to displaying the experimental results and performance metrics. The theoretical advances and computing efficiency attained by the optimization processes have immediate real-world applications in object tracking scenarios such as surveillance systems, traffic monitoring, and human activity analysis. Because of the durability exhibited in dealing with hard conditions such as occlusion and the unexpected appearance/disappearance of objects, this technology is particularly helpful for practical application in dynamic contexts. Enhanced measures such as accuracy, recall, mAP, and F1 score directly contribute to improving the dependability and efficacy of object tracking systems across multiple domains. The proposed methodology shows itself as a viable tool for real-world applications by stressing these practical consequences, addressing the demand for improved and efficient object tracking systems.

VI. Conclusion and Future Work

The combined strategy of using Particle Swarm Optimization (PSO), Kalman Filtering, Convolutional Neural Networks (CNNs), and Compressive Sensing has shown to be very successful in addressing the difficulties involved in multi-object tracking and identification in dynamic environments. By using CNNs for accurate object recognition, compressive sensing for well-tuned representations, Kalman Filtering for flexibility, and PSO for the best forecasting accuracy, this synergistic system performs better than individual techniques. The joint application of these methods showed excellent tracking performance under difficult conditions, such as complex motion patterns, occlusions, and visual discrepancies. This suggested methodology has notable practical benefits. With a tracking accuracy of 98%, the system demonstrates its potential for practical use, especially in fields that demand advanced multi-object recognition and tracking. In addition to advancing computer vision, the study establishes a strong basis for further advancements in multi-object tracking systems. It is important to recognize some restrictions, though. Real-time monitoring in resource-constrained contexts requires further optimization of the computational efficiency of the program. Investigating transfer learning and domain adaptation can improve the system's generalization in a variety of tracking environments. Three-dimensional object tracking would benefit from the addition of depth information, and more research is necessary to determine how robust the method is against inclement weather and abrupt changes in illumination. Future work can concentrate on optimizing the methodology for real-time application, increasing scalability to support a larger number of objects and a wider range of surroundings, and combining it with creative algorithms. The above efforts would certainly improve the capabilities of the methodology and increase its application across many areas. To sum up, our research offers significant understandings of the theoretical and practical aspects of multi-object tracking systems, opening new avenues for further investigation and development in computer vision. The suggested CNN-FCT methodology performs well in a variety of settings, but it is important to recognize its possible advantages as well as its limitations with regard to different kinds of data. To further improve adaptability and meet the particular obstacles presented by various circumstances, future study could investigate algorithm adaptations or parameter tailoring for certain data kinds.

References

[1] F. Schilling, F. Schiano, and D. Floreano, "Vision-Based Drone Flocking in Outdoor Environments," IEEE Robot. Autom. Lett., vol. 6, no. 2, pp. 2954–2961, Apr. 2021, doi: 10.1109/LRA.2021.3062298.

[2] A. A. Rafique, A. Jalal, and A. Ahmed, "Scene Understanding and Recognition: Statistical Segmented Model using Geometrical Features and Gaussian Naïve Bayes," in 2019 International Conference on

Applied and Engineering Mathematics (ICAEM), Taxila, Pakistan: IEEE, Aug. 2019, pp. 225–230. doi: 10.1109/ICAEM.2019.8853721.

[3] W. Ren, X. Wang, J. Tian, Y. Tang, and A. B. Chan, "Tracking-by-Counting: Using Network Flows on Crowd Density Maps for Tracking Multiple Targets," IEEE Trans. Image Process., vol. 30, pp. 1439–1452, 2021, doi: 10.1109/TIP.2020.3044219.

[4] A. Farid, F. Hussain, K. Khan, M. Shahzad, U. Khan, and Z. Mahmood, "A Fast and Accurate Real-Time Vehicle Detection Method Using Deep Learning for Unconstrained Environments," Appl. Sci., vol. 13, no. 5, p. 3059, Feb. 2023, doi: 10.3390/app13053059.

[5] J. Zhou, S. Zeng, and B. Zhang, "Linear Representation-Based Methods for Image Classification: A Survey," IEEE Access, vol. 8, pp. 216645–216670, 2020, doi: 10.1109/ACCESS.2020.3041154.

[6] N. Sindhwani, R. Anand, M. S., R. Shukla, M. Yadav, and V. Yadav, "Performance Analysis of Deep Neural Networks Using Computer Vision," EAI Endorsed Trans. Ind. Netw. Intell. Syst., vol. 8, no. 29, p. 171318, Nov. 2021, doi: 10.4108/eai.13-10-2021.171318.

[7] W.-G. Li and H. Wan, "An improved spatio-temporal context tracking algorithm based on scale correlation filter," Adv. Mech. Eng., vol. 11, no. 2, p. 168781401982590, Feb. 2019, doi: 10.1177/1687814019825903.

[8] C. R. Bern, K. Walton-Day, and D. L. Naftz, "Improved enrichment factor calculations through principal component analysis: Examples from soils near breccia pipe uranium mines, Arizona, USA," Environ. Pollut., vol. 248, pp. 90–100, May 2019, doi: 10.1016/j.envpol.2019.01.122.

[9] H. D. Najeeb and R. F. Ghani, "A Survey on Object Detection and Tracking in Soccer Videos," Muthanna J. Pure Sci., vol. 8, no. 1, pp. 1–13, Jan. 2021, doi: 10.52113/2/08.01.2021/1-13.

[10] Assistant Professor, Department of Computer Science and Engineering, IFET College of Engineering, Villupuram, India and Dr. D. S. David*, "An Intellectual Individual Performance Abnormality Discovery System i n Civic Surroundings," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 5, pp. 2196–2206, Mar. 2020, doi: 10.35940/ijitee.E2133.039520.

[11] M. M. Moussa and R. Shoitan, "Object-based video synopsis approach using particle swarm optimization," Signal Image Video Process., vol. 15, no. 4, pp. 761–768, Jun. 2021, doi: 10.1007/s11760-020-01794-1.

[12] N. Nedjah, A. V. Cardoso, Y. M. Tavares, L. D. M. Mourelle, B. B. Gupta, and V. Arya, "Co-Design Dedicated System for Efficient Object Tracking Using Swarm Intelligence-Oriented Search Strategies," Sensors, vol. 23, no. 13, p. 5881, Jun. 2023, doi: 10.3390/s23135881.

[13] X. Chen, F. Li, and X. Liu, "Efficient and Robust Distributed Digital Codec Framework for Jointly Sparse Correlated Signals," IEEE Access, vol. 7, pp. 77374–77386, 2019, doi: 10.1109/ACCESS.2019.2920982.

[14] D. T. Nguyen, T. N. Nguyen, H. Kim, and H.-J. Lee, "A High-Throughput and Power-Efficient FPGA Implementation of YOLO CNN for Object Detection," IEEE Trans. Very Large Scale Integr. VLSI Syst., vol. 27, no. 8, pp. 1861–1873, Aug. 2019, doi: 10.1109/TVLSI.2019.2905242.

[15] D. Li, Y. Yu, and X. Chen, "Object tracking framework with Siamese network and re-detection mechanism," EURASIP J. Wirel. Commun. Netw., vol. 2019, no. 1, p. 261, Dec. 2019, doi: 10.1186/s13638-0191579-x.

[16] S. Kaderali, "Detection and Characterisation of Unknown Spacecraft Maneuvers," 2021.

[17] I. Ullah, X. Su, J. Zhu, X. Zhang, D. Choi, and Z. Hou, "Evaluation of Localization by Extended Kalman Filter, Unscented Kalman Filter, and Particle Filter-Based Techniques," Wirel. Commun. Mob. Comput., vol. 2020, pp. 1–15, Oct. 2020, doi: 10.1155/2020/8898672.

[18] N. Mahmoudi, S. M. Ahadi, and M. Rahmati, "Multi-target tracking using CNN-based features: CNNMTT," Multimed. Tools Appl., vol. 78, no. 6, pp. 7077–7096, Mar. 2019, doi: 10.1007/s11042-018-6467-6.

[19] Y. Liu et al., "FedVision: An Online Visual Object Detection Platform Powered by Federated Learning," Proc. AAAI Conf. Artif. Intell., vol. 34, no. 08, pp. 13172–13179, Apr. 2020, doi: 10.1609/aaai.v34i08.7021.

[20] C. Li and B. Yang, "Adaptive Weighted CNN Features Integration for Correlation Filter Tracking," IEEE Access, vol. 7, pp. 76416–76427, 2019, doi: 10.1109/ACCESS.2019.2922494.

[21] P. Dendorfer et al., "MOTChallenge: A Benchmark for Single-Camera Multiple Target Tracking," Int. J. Comput. Vis., vol. 129, no. 4, pp. 845–881, Apr. 2021, doi: 10.1007/s11263-020-01393-0.

[22] "Papers with Code - MOT17 Dataset." https://paperswithcode.com/dataset/mot17 (accessed Aug. 14, 2023).

[23] L. Zhao and S. Li, "Object Detection Algorithm Based on Improved YOLOv3," Electronics, vol. 9, no. 3, p. 537, Mar. 2020, doi: 10.3390/electronics9030537.

[24] A. M. Roy, R. Bose, and J. Bhaduri, "A fast accurate fine-grain object detection model based on YOLOv4 deep neural network." arXiv, Oct. 30, 2021. Accessed: Jul. 18, 2023. [Online]. Available: http://arxiv.org/abs/2111.00298

[25] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs," Sensors, vol. 22, no. 2, p. 464, Jan. 2022, doi: 10.3390/s22020464.

[26] F. Klepel and R. Goebel, "Learning Equivariant Object Recognition and its Reverse Application to Imagery," Neuroscience, preprint, May 2023. doi: 10.1101/2023.05.20.541553.

[27] M. Wang, M. Lv, H. Liu, and Q. Li, "Mid-Infrared Sheep Segmentation in Highland Pastures Using Multi-Level Region Fusion OTSU Algorithm," Agriculture, vol. 13, no. 7, p. 1281, Jun. 2023, doi: 10.3390/agriculture13071281.