Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.18.(2) 2026,pp.Comp 99–113

University Of AL-Qadisiyah
Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

A Control-Oriented Information System for Low-Light Object Tracking Based on Multi-Level Image Enhancement and YOLOv12

Laith Hakem Malak [a*]

College of Computer Science and Information Technology, University of Al-Qadisiyah, Diwaniyah, Iraq .

Email: laithhakemalzeyadi@qu.edu.iq

ARTICLE INFO

Article history:
Received: 06 /04/2026
Rrevised form: 21 /04/2026
Accepted : 22 /04/2026
Available online: 30 /06/2026

Keywords:

Object detection, YOLOv12, Contrast Limited Adaptive Histogram Equalization , Lowlight image enhancement, tracking in harsh environment

ABSTRACT

Tracking and object detection in low-light conditions is one of the most complicated problems of computer vision, as the conditions of poor light, low contrast, and the disappearance of valuable structural image characteristics in photographed images make it difficult to detect objects. These are some of the restrictions that reduce the output of detection and tracking algorithms, especially in the cases of surveillance and monitoring. To deal with this issue, it is proposed in this paper that an integrated framework can be created to improve the enhancement of low-light images and object detection through a multi-level image fusion strategy in conjunction with the YOLOv12 detector model. The suggested approach will improve low-light frames by a multi-stage algorithm including enhancement of base illumination, amplification of the details layer, and amplification of contrast with the help of Contrast Limited Adaptive Histogram Equalization (CLAHE). Noise-sensitive mask is also included in such a way that the strength of improvements in the dark area can be controlled to avoid unnecessary increase in noise. The refined frames are then used by the YOLOv12 model as input to detect and track objects. A dataset of 79 images in seven groups that depict the scenes of different crowds was evaluated experimentally. Conditions of low light were artificially produced to mimic difficult light conditions. The performance of the enhancement was measured in Structural similarity index (SSIM) and peak signal to noise ratio (PSNR) and the performance of the detection process was measured in confidence scores and the number of detected objects. The findings of the experimental study show that the suggested framework contributes to the increase of the visual quality and object detection throughput during the low-light conditions, especially in dense crowd scenarios. The findings affirm the fact that the combination of image enhancement, combined with detection models, can be an effective means of enhancing tracking reliability and the visibility of the object in problematic settings.

https://doi.org/10.29304/jqcsm.2026.18.22788

1. Main text

Detecting and tracking objects have become critical in most of computer vision works such as intelligent surveillance systems, autonomous driving, traffic monitoring and security systems[1-3]. The success of these applications depends greatly on precise visual details to detect and trace the objects in a video series. As the sphere of deep learning evolves swiftly, the current object detectors like YOLO have contributed to a substantial increase in the quality of object detection and speed in the process in real-time[4].

∗Corresponding author

Email addresses:

Communicated by ‘sub etitor’

Despite these developments, object detection and tracking systems continue to face significant challenges in low-light environments [5]. Videos taken in low light environments are usually characterized by low brightness, low contrast, and lack of the structural details. The degradations make objects less visible and events that occur greatly influence the output of detection and tracking algorithms. Under most real-world scenarios like the nighttime video monitoring, monitoring underground or insufficiently lit abodes of the interior, the video information or data obtained may fail to generate reliable visual data that one can use to track objects towards the desired sense of surety [6, 7].

To deal with this issue, an image enhancement mechanism is in most cases used to enhance the visual quality of low-light images prior to conducting any high-level vision functions [8, 9]. The detection and tracking models can function on clearer images of what they see, which increases the precision and stability of the detection and tracking models [10].

1.1. Motivation

Object detection and tracking algorithms mainly rely on the quality of input visual information. In the low-light environments in which the video is being shot, the deteriorated image quality may result in the inappropriate recognition of objects, unreliable tracking, and targeted loss. Conventionally used object tracking techniques usually suppose that the input frames have adequate visual information that are not always available in the real world setting [5, 6].

Despite the fact that a number of low-light image enhancement techniques have been suggested in the recent past, most of these techniques are motivated by improving visual characteristics of an image despite the effects that they will have on later computer vision algorithms like object detection and tracking [11]. Moreover, there are also enhancement methods that can add noise amplification or unnatural brightness intensities that can harm the tracking algorithm's performance [5].

Thus, it should be seen that there is a necessity of an integrated framework, which does not only contribute to better images during low-light, but the stability of the object detection and tracking systems. Driven by this problem, this paper presents a multi-level video enhancement framework, which can be used to enhance low-light video images without compromising structural information and balancing noise, thus making object detection, and tracking more dependable[12, 13].

1.2. Contributions

The principal contributions of this work may be described as follows:

1.A multi-level image enhancement system under low-light conditions is suggested to enhance brightness, contrast, and structure details of low-light video frames.

2..An appropriate enhancement plan is provided and involves noise control mask where the intensity of the improvement in faulty areas is controlled to avoid amplification of noise.

3.It comes up with a fusion-based decision-making framework that combines base illumination-enhancement, detail-layer-enhancement as well as contrast-enhancement with adaptive fusion weights.

4.A processing pipeline will be constructed based on the detection and tracking of objects with the help of the YOLOv12 model, as well as low-light processing.

5.The suggested framework enhances the stability and strength of tracking objects in the dark environment by supplying the detection and tracking model with visually enhanced frames.

The proposed model constitutes the single system that increases the video in low-light conditions and includes an object detection system and tracking on a video in one processing stream. The architecture takes a low-light video stream which exits a camera apart into elementary components. frames. The next processing is done with a multi-level enhancement algorithm which enhances the illumination and the contrast and maintains valuable structure details of the image.

The enhancement process works using a number of visual elements of the picture such as bottom illumination layer, the details layer, and the contrast data [8]. These elements are boosted individually, and then mixed with each other with a weights-based fusion mechanism governed by a noise-sensitive mask. It is a strategy that enables gradual enhancements of brightness and still preserves a natural look and avoids too much noise enhancement [14].

The refined frames are given to the YOLOv12 model after the enhancement phase to detect and track objects. The proposed framework enhances the tracking system by the capability of every succeeding frame to identify and track objects accurately by detecting and tracking on enhanced frames as opposed to the original low-light frames.

Lastly, the reconstructed processed frames are then reconstituted back into a video sequence consisting not only of improved visual quality but also of trustworthy tracking outputs. The suggested solution is thus quite useful in enhancing the object detection and tracking performance in low-light environments[15].

Despite the numerous studies conducted to investigate the process of enhancing low-light images and detecting objects, there are still a number of challenges to address. However, most existing methods focus on improving visual quality without explicitly considering their impact on object detection and tracking performance in low-light environments[2, 16]. Also, enhancement models of deep learning usually demand vast amounts of data and massive computation. In addition, numerous detection-based methods alter the model of detection itself but do not deal with the quality of the given picture.

To overcome these drawbacks, this work suggests a multi-level low-light improvement system with embedded YOLO-based object tracking and detection, to assist the quality of visibility and the stability of tracking in the dark video clip.

2. Related works

computer vision due to its importance in applications such as surveillance systems, autonomous vehicles, and intelligent monitoring systems. However, images captured under poor illumination conditions often suffer from low brightness, high noise levels, and reduced contrast, which negatively affect the performance of high-level vision tasks such as object detection and tracking. Researchers have proposed various methods to address these challenges, which can generally be categorized into traditional enhancement methods, deep learning-based enhancement approaches, and integrated detection-enhancement frameworks.

Although these methods improve low-light image quality or detection performance, most of them do not provide an integrated framework that jointly optimizes enhancement and object detection for improved robustness in low-light environments.

Mohammed et al. introduced the well-known Contrast Limited Adaptive Histogram Equalization (CLAHE) technique, which enhances local contrast in images while limiting noise amplification. CLAHE operates on small image regions and redistributes pixel intensities locally, making it particularly useful for improving visibility in dark regions [17].

Jobson et al. proposed the Retinex theory-based enhancement method, which simulates human visual perception by separating illumination from reflectance components. Retinex-based approaches have been widely used to improve brightness and dynamic range in low-light images [18].

Wang et al. provided a comprehensive review of classical low-light enhancement methods, including histogram equalization, gamma correction, and Retinex-based techniques. Their study highlighted that traditional methods are computationally efficient but often suffer from problems such as over-enhancement and noise amplification [19].

Guo et al. proposed Zero-DCE, a deep curve estimation network for low-light image enhancement. The model learns a set of pixel-wise adjustment curves to improve illumination without requiring paired training data. The method demonstrated strong performance with low computational complexity[11].

He et al. introduced an attention-based network for low-light image enhancement, which utilizes spatial and channel attention mechanisms to selectively enhance important visual features while suppressing noise in dark regions [20].

Jiang et al. proposed an unsupervised low-light image enhancement framework that decouples illumination enhancement and noise suppression tasks. This architecture allows the network to independently learn these components and achieve better enhancement performance [21].

Peng et al. proposed NLE-YOLO, a low-light target detection network based on YOLOv5. Their model integrates feature enhancement modules and adaptive learning strategies to improve object detection accuracy in low-light environments [22].

Liu et al. introduced Dark-YOLO, a modified YOLO-based object detection algorithm designed for low-light scenes. The model improves feature extraction and enhances detection accuracy by incorporating illumination-aware modules[23].

Han et al. proposed 3L-YOLO, a lightweight object detection model optimized for low-light environments. The model focuses on improving feature representation while maintaining computational efficiency for real-time applications [24].

3. Material and methods

The section gives the theoretical concepts on which the proposed framework is based. Such topics are low-light image degradation, image enhancement methods, contrast enhancement, and object detection approaches based on deep learning models, e.g., YOLO.

3.1. Low-Light Image Degradation

Low-light photographs also tend to be affected by some of the degradation factors that include low brightness, low contrast, and high noise. Such problems cause the lack of illumination in image sensors, which would lead to a low signal-to-noise ratio (SNR)[25]. This makes crucial structural information and object outlines hard to be differentiated with the background. Overall, the captured low-light image can be represented as the composition of the illumination component and the reflectance component. The illumination-reflectance model would allow an image, I(x), to be modeled as shown in equation 1.

$$I(x)=R(x)\times L(x) \qquad (1)$$

where I(x) represents the observed image intensity, R(x) represents the reflectance component corresponding to the intrinsic properties of objects, and L(x) denotes the illumination component. In low-light conditions, the illumination component becomes very weak, which leads to dark images with reduced visibility.

The goal of low-light enhancement methods is to recover or improve the illumination component while preserving the reflectance information that contains important structural details of the scene.

3.2. Image Enhancement Techniques

Image enhancement technologies are designed to execute the enhancement of visual quality of pictures, enhancing the brightness and contrast of images, as well as their structural features [26]. Such techniques can broadly be classified into spatial domain techniques and frequency domain techniques. The spatial domain techniques make direct manipulations of pixel values to enhance the quality of an image. Popular methods are gamma correction, histogram equalization, and methods based on the Retinex [27]. A gamma correction process controls or adjusts the brightness of an image by applying nonlinear transformations to the pixel value of the image, whereas histogram equalization rebalances pixel values of an image to increase contrast. The retinex-based methods are common in low-light boosting since they replicate the human eye. The techniques seek to isolate the illumination and reflectance components and boost the illumination component to add to the overall brightness without losing details [8]. The conventional methods of enhancement might, however, add unwanted artifacts like over-enhancement or noise amplification, especially in very dark areas. Hence, aggressive performance enhancement plans are usually necessitated in sophisticated performance. Contrast enhancement is significant in enhancing the visibility of objects in low-light images[28]. Contrast Limited Adaptive Histogram Equalization (CLAHE) is one of the techniques widely used.

In comparison with global histogram equalization, CLAHE works on small regions of the image known as tiles. Histograms of both areas are also equalized separately thereby enhancing the local contrast as well as bringing out hidden features in dark regions [27]. A clipping threshold is used to store the amplitude to avoid a huge amplification of noise. CLAHE boosts the quality of visibility of low-light images by boosting local contrast at the expense of noise and facilitates the recognition of meaningful structures according to the later computer vision algorithm.

3.3. Object Detection in Computer Vision

Object detection is a core operation of computer vision to detect and localize objects in an image or in a video frame [5]. It is aimed at identifying the type of the object as well as the spatial position of the object with the help of bounding boxes [6]. Deep techniques of the contemporary object detection rely mainly on convolutional neural networks (CNNs). These models are auto hierarchical feature learners that can find objects with great accuracy on large collections of data [22].

Object detection algorithms have broadly been divided into two major groups:

1.Two-stage instruments, including R-CNN and Faster R-CNN, that initially create region recommendations and then recognize them.

2. Single-stage detectors e.g. YOLO (You Only Look Once) and SSD, which do the detection and classification in a single forward pass.

Due to the speed and the high computational efficiency, single-stage detectors find many applications in real-time. One of the most popular object detection frameworks in real-time is YOLO (You Only Look Once). YOLO is an object detector that solves a regression problem directly on the input image and predicts bounding boxes and probabilities of different classes: it is a single-pass network that performs object detection.[5]

The YOLO model divides the input image into a series of grid cells and perceives a series of bounding boxes on each of these grid cells. Each predicted bounding box contains the object localization, likelihood rating, and probability of class. This single-stage method of detection helps the YOLO models to attain a high rate of detection and, at the same time, be competitive in terms of accuracy [4].

Recent extensions of YOLO have come up with several improvements in feature extraction, multi-scale detection, and training. The improvements enable the YOLO-based models to work well in demanding settings, including those with low light.

Object tracking is designed to trace the path of the identified objects in succession through a video sequence. Tracking algorithms infer the location of the previously identified objects in the later frames through motion information and appearance measures after saving data in the first frame.

Tracking of objects properly depends on the quality of the input frame a lot[29]. When the light is low, tracking algorithms may cause tracking errors such as missing the target and poor contrast leading to inaccuracy in the tracking result. Hence, making the video frames have a better visual appearance prior to object detection and tracking may be of great help in increasing the tracking stability and reliability significantly.

3.4. Proposed model

The suggested model will present a combined system that fuses both low-light video enhancement and object tracking in low-light conditions. This framework is primarily aimed at enhancing the visual quality of the video frame in low-light conditions prior to the tracking set. The tracking algorithm can be used with the cleaner visual data produced by the process of increasing the brightness and contrast of the frames, leading to an increase in the reliability and stability of the object tracking.

In most live situations, videos taken in the dark suffer poor brightness, contrast and structural details. The effects of these limitations are much on the way the object tracking algorithms perform since at this point, the targets will be hard to be differentiated between the background. The proposed model is expected to overcome this limitation by combining an image enhancement step with a tracking step in a single processing pipeline.

The processing structure starts with an input of a low-light video stream, which is captured by a camera. The video input is initially broken down into a series of different frames. The proposed enhancement method is then applied to each frame individually with the purpose of enhancing the real brightness, increasing the contrast, and saving the valuable structural data. This increase action seeks to create frames whose visual features have more articulateness and minimize the adverse impacts of lack of light.

Object detection and tracking are also done by feeding the improved frames to the YOLOv12 model after the enhancement stage has been successfully completed. With the tracking process applied on the enhanced frames instead of the initial low-light frames, the proposed performance framework can be used to ensure the tracking model detects the targets better, and the tracking performance is also stable between adjacent frames.

In the suggested solution, the enhancement of a low-light picture is achieved by applying several levels of enhanced illumination. Rather than using one global operation to process the image, multiple steps are used with each level processing a certain visual quality, like illumination, contrast, or the detail of the structure.

The principal proposed model can be found in Figure 1

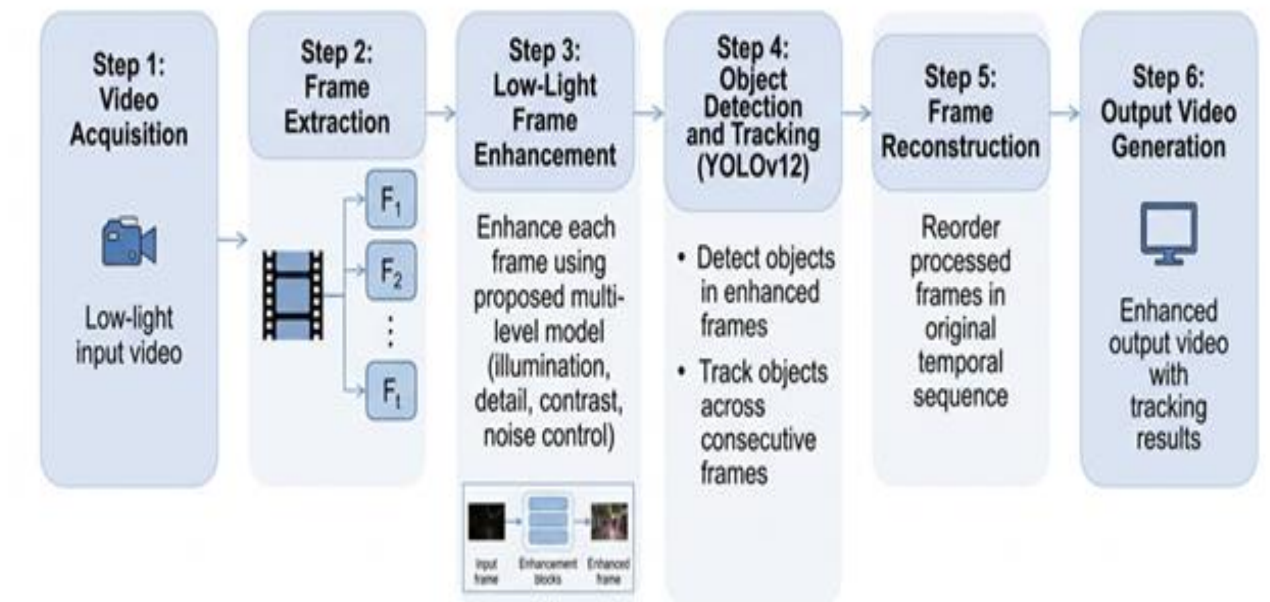


Fig 1: proposed model

Step 1: Video Acquisition

It starts with taking a low-light video stream of a camera. In bad lighting conditions, the video being taken is likely to have dark spots, low contrast and less visibility of the object. This unedited video is the input of the suggested framework.

Step 2: Frame Extraction
The input video is separated into a series of single frames. Each frame is an extract static image out of the video at a particular time interval. The step enables the process of enhancement to be extended separately to each frame.

Step 3: Low-Light Frame Enhancement

In this step, every frame is pulled out is enhanced by the suggested low-light enhancement technique. Improvement of light, enhancement of contrast, and retention of structural details of the objects in the frame are the benefits of the process of enhancement. The step produces better and visual frames that are better placed to detect and track objects.

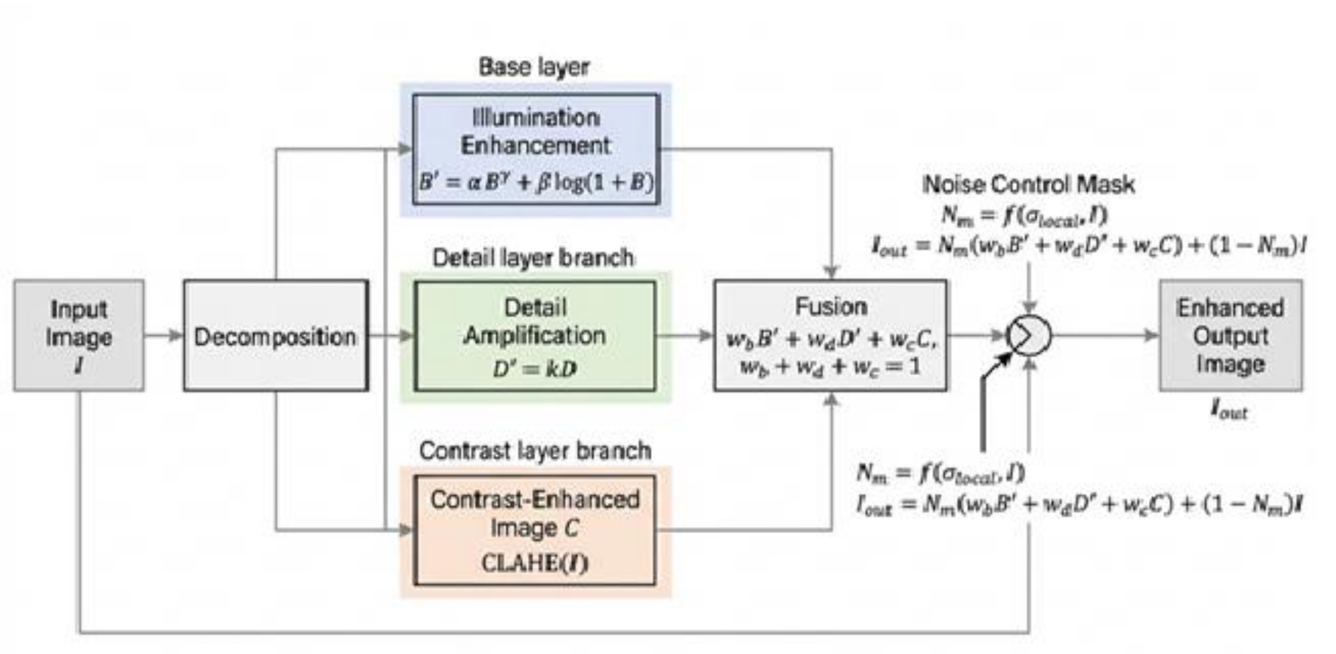


Fig 2: Proposed low light image enhancement

This is a multi-level approach that has the ability to enhance the brightness of the image step by step without the unnaturally high brightness but without the problematic amplification of noises.

The procedure starts with the dissection of an input image into differing fundamental elements. Mathematical transformations are then applied to each of the components and finally weighted fusion is used to recombine all the enhanced components.

The above fusion model can be used to describe the overall enhancement process.:

$$I_{\text{out}} = N_m(w_b B' + w_d D' + w_c C) + (1 - N_m)I \quad (2)$$

Where:

I_{out} denotes the final enhanced output image. N_m represents the noise control mask that regulates the enhancement strength in noisy regions.

w_b, w_d , w_c are fusion weights satisfying w_b denotes the fusion weight of the enhanced base illumination layer. w_d denotes the fusion weight of the enhanced detail layer. w_c denotes the fusion weight of the contrast-enhanced image. The $w_b + w_d + w_c = 1$ B' represents the enhanced base illumination component. D' represents the amplified detail component. C denotes the contrast-enhanced image. I represents the original input image. Nm is in charge of the effect of the enhancement process. To minimize the strengths of enhancement in the higher noise levels, Nm reduces whereas, to enhance the strengths of enhancement by the processed layers, Nm is given extra significance in order to avoid excessive amplification of noise in extremely dark areas. A noise control mask is added to curb the excessive amplification of noise at extremely dark areas. This mask balances the intensity of the enhancement to the local noise present in an image. The noise mask is defined as shown in equation 3:

$$Nm = f(\sigma_{local}, I) \quad (3)$$

where:

N_m denotes the noise control mask.

σ_{local} represents the local standard deviation of pixel intensities computed within a small neighborhood window. It is used as an estimate of local noise or texture strength.

I denotes the original input image.

$f(\cdot)$ is a mapping function that generates a weighting value between 0 and 1 .

The value of N_m determines the contribution of the enhanced layers in the final fusion stage. Regions with high local variance (which may indicate noise in low-light conditions) receive smaller mask values, reducing enhancement strength. Conversely, regions with stat ⇓ tensity variations receive higher mask values, allowing stronger enhancement.

The σ_{local} calculated in equation 2x

$$\sigma_{\text{local}}(x,y) = \sqrt{\frac{1}{N}\sum_{(i,j)\in\Omega} (I(i,j)-\mu_{\Omega})^2} \qquad 2x$$

where Ω represents a local neighborhood around pixel (x,y), N is the number of pixels within that region, and μ_{Ω} is the mean intensity in the neighborhood. This measure can be used to estimate the presence of texture or noise in different regions of the image.

In many enhancement frameworks, multiple enhanced components are combined to produce the final image. A general image fusion model can be expressed as

Level 1: Base Illumination Enhancement

A nonlinear transformation of base layer is added to improve the latter based on the combination of power-law correction and logarithmic adjustment. A nonlinear illumination function is used to add on to the base layer.:

$$B' = \alpha B^{\gamma} + \beta \log(1+B)$$

Where:

α = brightness scaling factor
γ = illumination correction factor
β = logarithmic enhancement factor

Level 2: Detail Layer Amplification

The detail layer preserves edges and fine structures:

D=I−B

It is amplified using a detail factor:

D′=kD

Where:

k= detail gain factor

Layer 3: Contrast Enhancement Layer

At this stage, another contrast-enhanced image is created in attempt to make the image more visible in dark areas, and enhance the local intensity values. In this work, contrast layer is realized with the help of Contrast Limited Adaptive Histogram Equalization (CLAHE) which is applied to the input image..

The contrast-enhanced image is defined as

C=CLAHE(I)

where:

C denotes the contrast-enhanced image.

I represents the original input image.

CLAHE is an adaptive histogram equalization method that improves local contrast while limiting noise amplification through a clipping threshold.

Through small image regions (tiles), in comparison to global histogram equalization, CLAHE redistributes the intensity values there confined regions. This gives the ability to boost dark areas and maintain natural brightness and not to over saturate bright areas.

This contrast layer is used to add further structure details which are providing later in the final step of enhancement which is fused with the illumination and detail layer.

Step 4: Object Detection and Tracking

Once the process of enhancing frames is done, the enhanced ones are fed to the YOLOv12 model. Object detection and tracking are done in the model, and this involves determining the target in every frame and tracking its location in subsequent frames. Given the fact that the frames already are improved, the tracking model is able to identify objects more distinctly and preserve a high level of tracking accuracy.

Step 5: Frame Reconstruction

At the end of the tracking process, the processed frames are reassembled as per their original time sequence. This step reassembles the video sequence at the same time keeping results of tracking on the improved frames.

Step 6: Output Video Generation

The final product of the strategy is the creation of a recreated video whereby the augmented visual precision and the tracking results are simultaneously symbolized. By combining the lighting intensity support and the tracking of the objects in a unified system, the developed model enhances the effectiveness of the tracking algorithm, which allows it to work effectively in the dark environment.

4. Experimental Results and Evaluation

4.1. Dataset Preparation

As the performance of the suggested framework was to be assessed, a series of video frames with crowds pictures and people assemblies were utilized. Rather than relating to the complete video sequences, frames were extracted and subsequently used as single images to be experimented on in a controlled world. Each frame is a scene with one or more people situated in a congested setting.

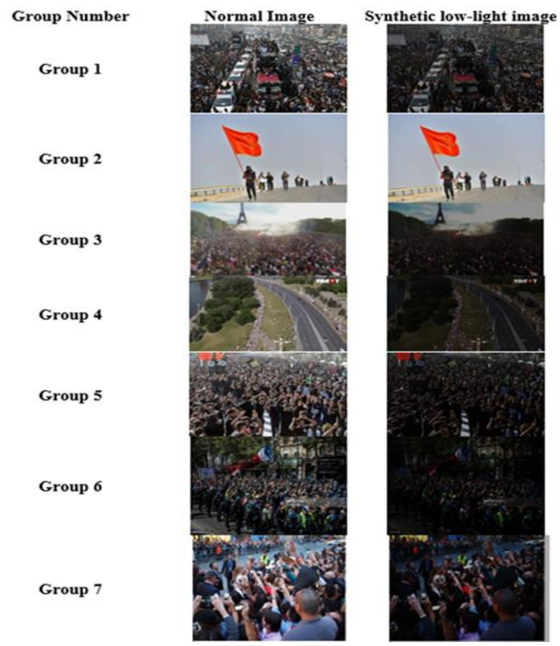
The chosen frames consist of different levels of density of the crowd, such as small groups of people up to a very high density of the crowd. These pictures were employed to model the real life surveillance situation where object detectors and trackers are commonly implemented.

To test how robust the proposed model will be in the illumination conditions that are considered adverse, artificial low-light conditions of the original images were enforced. The contrast and brightness of the frames were deliberately lowered in order to replicate poor lighting conditions in surveillance at night and poorly lit indoors. The approach enabled the test of the proposed method of enhancements to be controlled; however, the original structure of the scene was retained.

The proposed framework was fed with the artificially degraded images. All the pictures were pre-enhanced using the proposed low-light improvement algorithm to enhance illuminance and contrast without distorting significant structural features of the subjects in the image. Following the transformation of the enhancement stage, the provided enhanced frames were sent to the object detection and tracking model to estimate how effective focus on the enhancement phase to the detection performance is.

The experiment setting can be used to assess the capability of the proposed model in improving the quality of visual information and increasing the robustness of object detection using frames of a crowd scene with artificially generated low-light conditions.

It is also critical to demonstrate the examples of the images that were used in the experiments before introducing the experimental results. The data applied in this paper is comprised of 79 photos followed into seven categories with each category representing varied scenes of the crowd with varied numbers of individuals. This was done using these pictures in the form of separate frames to test how well the proposed model would perform in low-light situations. As Figure 3 indicates, one of the crowd scenes of the dataset used in the experiments is presented in a sample image.



4.2. Image enhancement

In quantitative assessment of the effectiveness of the proposed low-light enhancement model, two commonly used image quality measures were used, namely Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). SSIM is used to quantify the structural similarity of the enhanced image and the reference image and represents the extent to which structural details are maintained. Conversely, PSNR measures the quality of reconstruction of the improved image compared to noise reduction and signal fidelity.

Table 1 shows the mean values of SSIM and PSNR of the seven groups of datasets used in the experiments.

The findings demonstrate that the enhanced model is capable of having different degrees of performance in various groups of images under varying complexity of scenes and intensity of crowds. Group6 was the most popular group among the ones described, with an SSIM value of 0.92 in addition to a PSNR value of 0.39, which is to show that the proposed approach is successful in its task of preserving a lot of structural data, as well as ensuring a high quality of reconstruction in this group. This finding implies that the enhancement system is especially effective in the scenes where there is a strong tendency of structural elements that can be easily noticed even when there is a low level of illumination.

Similarly, Group4 and Group5 also demonstrated strong performance, achieving SSIM values of 0.80 and 0.79, respectively, with PSNR values of 0.38 and 0.35. These results indicate that the proposed enhancement method is capable of significantly improving visual quality in moderately complex crowd scenes while preserving important structural details.

On the other hand, Group2 recorded the fewest SSIM (0.52) and comparatively smaller PSNR of 0.28. It could be easy to explain this poorer performance by the complexity of the scene that was raised or the noise that was added by the low-light conditions that had been simulated. Under these conditions, differentiating structural features is not as straightforward, which will have a slight influence on the process of improvement.

Other groups (Group1, Group3, and Group7) obtained moderate performance with the values of SSIM between 0.66 and 0.75 as well as PSNR between 0.26 and 0.33. These findings show that the developed algorithm is reliable to enhance the quality of images in various scene setups and at the same time retain decent structural similarity.

Altogether, the quantitative findings indicate that the developed multi-level improvement model is useful in enhancing the image quality under low-light conditions. The observed relatively high SSIM values in most conditions imply that the structural information of the original scenes is not deteriorated, whereas the outcome of the process of PSNR measurement knows that the process of enhancement preserves the acceptable signal fidelity without bringing a lot of noise.

All these findings indicate that the proposed strategy to enhance the visual quality in low-light crowd scenes is appropriate and generates dependable input to be used in the tasks of object detection and tracking afterwards.

Figure X provides examples of the dataset used in the experiments that are qualitative in nature. The figure demonstrates sample pictures of the seven groups that were incorporated in the dataset. Each group will be presented with three variants of the same image: the initial normal image, the changed image.

synthetic low-light image, and the enhanced image of the proposed method.

The synthesis of the low light images used in the synthetic low light images was done through the artificially diminishing of the brightness and contrast of original images to replicate the poor illumination conditions as in the figure. Such low-quality images are a depiction of tricky situations that are normally experienced in surveillance platforms where one might be working at night or in poorly lit areas.

The third column demonstrates the improved findings with the help of the suggested low-light enhancement framework. As it could be seen, the suggested process is really effective to enhance the total light and revive significant visual details that could not be distinctly seen in the low-light pictures. Specifically, the increased brightness, contrast and clarity of structural information of the scene has been exhibited in the enriched images, particularly in the places where there are high numbers of people.

Additionally, the process of enhancement does not destroy the natural look of the images and does not solely increase the noise overload. Such an increase in image quality gives better input information in the next steps of object detection and tracking of the proposed structure.

Comprehensively, the qualitative findings introduced by Figure X provide the evidence of the feasibility of the proposed model to be successfully applied in improving low-light images in various case scenarios and intricacies of the scene.

Table 1: Evolution of proposed image enhancement model

group	SSIM	PSNR
Group1	0.75	0.33
Group2	0.52	0.28
Group3	0.66	0.26
Group4	0.80	0.38
Group5	0.79	0.35
Group6	0.92	0.39
Group7	0.72	0.32

4.3. Object detection

Comparison has been done between the results of detection prior to enhancement and the results obtained after the application of the proposed low-light enhancement model. The analysis uses two critical metrics to detect the confidence of detection and the count of objects found on each of the seven dataset groups.

As can be seen in Table X, there is an observable positive response to the proposed enhancement framework as far as detection performance is concerned. Generally, several objects have been detected in a larger number of groups which is an indicator that the improved images present better visual data to the detection model.

As an example, Group1 has been provided where the individuals who have been detected have risen to 17, whereas the confidence score has risen to 0.92 as compared to 11, 0.65 respectively. This shows that there occurred the improvement of detection accuracy and the object visibility by the enhanced process.

In the same way, Group2 demonstrates a great improvement, in terms of which the confidence score rose by 0.320.96, and the number of identified objects rose by 1220. This proves that the suggested improvement technique is especially efficient in the scenes that had a very low score of illumination, and the initial detection level was constrained.

In Group3, an even greater increase may be noted as the objects identified rose to 85 and the confidence rate rose by 0.08, IBM. This outcome brings out the potential of the proposed approach in high-crowd scenarios where the visibility of objects is critical in detection accuracy.

Group4 Group4 showed improvement in the number of objects detected in the post-enhancement images as compared to 10 items used in Group4, and most of the objects not detected in the original low-light images were detected during the post-enhancement images.

The Group5 results also indicate a major enhancement as the number of objects detected also went up by 16 (19 to 29) with the confidence also showing a moderate growth by 0.68 (0.79).

In case of Group6, despite a slight reduction in the score of confidence, 0.85 to 0.56, the occurrence of objects detected in the scene increased exponentially, which was 39 to 125 indicating that the enhancement method revealed a large number of the objects present in the scene that remained hidden previously. This fact demonstrates that the suggested approach enhances the visibility of objects when a crowd is especially large.

Last but not least, the Group7 also demonstrates better performance in which the detection count has risen to 56, and the confidence score has risen to 0.89 instead of 23 and 0.80 respectively.

The general findings of the experiment indicate that the presented low-light enhancement framework can substantially enhance the work of object detectors by removing either increasing the number of detected objects or by enhancing the detection confidence of the object detectors in the vast majority of scenarios. Such results validate the fact that the algorithm of augmentation of low-light images, prior to object detector application, can significantly increase the accuracy of detection structures, especially when the objects of interest are very close together.

Table 2: Compere of effects of proposed model on YOLO 12

	Before		After	
group	Confidence	count	Confidence	count
Group1	0.65	11	0.92	17
Group2	0.32	12	0.96	20
Group3	0.92	29	0.98	85
Group4	0.73	10	82	36
Group5	0.68	9	0.79	29
Group6	0.85	39	0.56	125
Group7	0.80	23	0.89	56

The visual comparison of results of object detection prior to implementation of the proposed low-light enhancement framework, and subsequent to the implementation, is displayed in Figure X. The figure shows a number of congested images in which the detection model locates the individuals in terms of bounding boxes and confidence scores.

The results of the detection before enhancement in the images displayed in the first row show that the model can detect only a few persons because of inappropriate illumination and lower visibility of objects. Certain individuals are either missed or identified with fairly low scores in confidence. This is largely limited by the fact that the quality of the images captured in low light is poor and thus it may be hard to clearly differentiate the features of objects as required by the detection model.

The performance of the identification is much better after the enhancement technique using the proposed method as it can be seen in the lower pictures. Images being enhanced display increased structural information and clear outlines of objects allowing the detection model to determine more individuals in the crowded scenes. Consequently, additional scenes of bounding boxes can be seen in the boosted images, which means that the number of detected objects increases.

Also, the increase in illumination and contrast assists the detection model to come up with more dependable predictions, which leads to the increased detection counts in the quantitative findings. These qualitative results prove that the suggested enhancement system is effective to enhance the visibility of items in the low-light space and facilitate more precise object detection when operating in the conditions of crowded crowds.

Fig 4:result of object tracking

Conclusion

In this paper, a combination of object detection and tracking enhancement in low-light conditions has been introduced in a comprehensive manner. The offered solution is a hybrid of a multi-level low-light image-enhancing procedure and a YOLOv12-based detection and tracking framework. The primary goal of the framework is to improve the visual quality of video frames that are taken in an unfavorable light environment, and then the object detection and tracking algorithms are applied.

The enhancement method works on various processing steps, such as base illumination enhancement, detail layer amplification process, and contrast enhancement with the help of CLAHE. Furthermore, it also features a noise-conscious mask that enables the control of the strength of the enhancement and eliminates noise upsurge in the very dark areas. The method combines these components in a weighted fusion mechanism, which allows brightness and contrast enhancement at the expense of valuable structure of the scene.

The experiment was evaluated with the help of the dataset, which contained 79 pictures, which were grouped into seven categories that reflected various crowd situations. Simulation of challenging illumination conditions was undertaken by the use of artificial low-light conditions. The experimental results of the quantitative assessment of the proposed method of enhancement based on the SSIM and PSNR measures proved that the new method is effective in enhancing the quality of the image and preserving the structural similarity. Additionally, the detection performance indicated that there was a vast enhancement in objects to be detected and the confidence of detection on the proposed improvement framework.

The qualitative findings also affirms that the improved images are clearer to give out visual information, making the model of detection to identify people with more accuracy in crowded scenes. In general, the experimental findings show that low-light image enhancement and object detection/tracking can significantly enhance the reliability of detection and functioning of the system in a low-light setting.

Further research on the presented framework can be termed as adding deep learning-based enhancement models and experimenting with the system on larger real-world video datasets in the future. As well, additional innovations can be made that will optimize the improvement parameters and incorporate more advanced tracking algorithms to increase the performance in dynamically changing scenes.

References

[1] G. Alhamzawi, A. A. Saeed, S. M. Hadi, A. H. Alsaeedi, A. A. Ahmed, M. R. Hassan, N. S. M. Satar, and W. Y. Yasseen, "Tuning adaptive gamma correction (TAGC) for enhancing images in low light," in *2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS)*, 2025: IEEE, pp. 1-6.

[2] Z. Jingchun, G. Eg Su, and M. Shahrizal Sunar, "Low-light image enhancement: A comprehensive review on methods, datasets and evaluation metrics," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 10, p. 102234, 2024/12/01/ 2024, doi: https://doi.org/10.1016/j.jksuci.2024.102234.

[3] X. Yi, H. Xu, H. Zhang, L. Tang, and J. Ma, "Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 12302-12311.

[4] D. Nimma, O. Al-Omari, R. Pradhan, Z. Ulmas, R. V. V. Krishna, T. Y. A. B. El-Ebiary, and V. S. Rao, "Object detection in real-time video surveillance using attention based transformer-YOLOv8 model," *Alexandria Engineering Journal*, vol. 118, pp. 482-495, 2025/04/01/ 2025, doi: https://doi.org/10.1016/j.aej.2025.01.032.

[5] S. Abba, A. M. Bizi, J.-A. Lee, S. Bakouri, and M. L. Crespo, "Real-time object detection, tracking, and monitoring framework for security surveillance systems," *Heliyon*, vol. 10, no. 15, p. e34922, 2024/08/15/ 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e34922.

[6] A. Elhanashi and S. Saponara, "Introduction to Object Detection and Localization," in *Deep Learning for Object Detection and Localization*, A. Elhanashi and S. Saponara Eds. Singapore: Springer Nature Singapore, 2026, pp. 1-24.

[7] X. Xu, R. Wang, C.-W. Fu, and J. Jia, "Snr-aware low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 17714-17724.

[8] L. He, Z. Yi, C. Chen, M. Lu, Y. Zou, and P. Li, "Detail-preserving noise suppression post-processing for low-light image enhancement," *Displays*, vol. 83, p. 102738, 2024/07/01/ 2024, doi: https://doi.org/10.1016/j.displa.2024.102738.

[9] H. Zhao, S. Xu, L. Peng, H. Hu, and S. Jiang, "Efficient Gamma-Based Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," *Applied Sciences*, vol. 15, no. 13, p. 7382, 2025. [Online]. Available: https://www.mdpi.com/2076-3417/15/13/7382.

[10] A. H. Alsaeedi, S. M. Hadi, and Y. Alazzawi, "Adaptive Gamma and Color Correction for Enhancing Low-Light Images," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 4, pp. 188-201, 2024.

[11] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong, "Zero-reference deep curve estimation for low-light image enhancement," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 1780-1789.

[12] R. R. Nuiaa, S. Manickam, and A. H. Alsaeedi, "Distributed reflection denial of service attack: A critical review," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6, p. 5327, 2021.

[13] R. R. Nuiaa, S. Manickam, A. H. Alsaeedi, and D. E. J. Al-Shammary, "Evolving Dynamic Fuzzy Clustering (EDFC) to Enhance DRDoS_DNS Attacks Detection Mechnism," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 1, pp. 509-519, 2022.

[14] J. Zhao, X. Wang, C. Zhang, J. Hu, J. Wan, L. Cheng, S. Shi, and X. Zhu, "Urban Waterlogging Monitoring and Recognition in Low-Light Scenarios Using Surveillance Videos and Deep Learning," *Water*, vol. 17, no. 5, p. 707, 2025. [Online]. Available: https://www.mdpi.com/2073-4441/17/5/707.

[15] M. Radif, A. H. Al-saeedi, R. R. Nuiaa Alogaili, and M. A. Mohammed, "Enhancing Classification Accuracy through Cluster-Based Ensemble Learning and Adaptive Weighting," *Fusion: Practice & Applications*, vol. 21, no. 1, 2026.

[16] C. Lee, C. Lee, and C. S. Kim, "Contrast enhancement based on layered difference representation," in *2012 19th IEEE International Conference on Image Processing*, 30 Sept.-3 Oct. 2012 2012, pp. 965-968, doi: 10.1109/ICIP.2012.6467022.

[17] I. M. Mohammed and N. A. M. Isa, "Contrast limited adaptive local histogram equalization method for poor contrast image enhancement," *IEEE Access*, 2025.

[18] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965-976, 1997, doi: 10.1109/83.597272.

[19] Y. Wang, R. Wan, W. Yang, H. Li, L.-P. Chau, and A. Kot, "Low-light image enhancement with normalizing flow," in *Proceedings of the AAAI conference on artificial intelligence*, 2022, vol. 36, no. 3, pp. 2604-2612.

[20] Z. He, W. Ran, S. Liu, K. Li, J. Lu, C. Xie, Y. Liu, and H. Lu, "Low-light image enhancement with multi-scale attention and frequency-domain optimization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 4, pp. 2861-2875, 2023.

[21] Q. Jiang, Y. Mao, R. Cong, W. Ren, C. Huang, and F. Shao, "Unsupervised decomposition and correction network for low-light image enhancement," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19440-19455, 2022.

[22] D. Peng, W. Ding, and T. Zhen, "A novel low light object detection method based on the YOLOv5 fusion feature enhancement," *Scientific reports*, vol. 14, no. 1, p. 4486, 2024.

[23] Y. Liu, S. Li, L. Zhou, H. Liu, and Z. Li, "Dark-Yolo: A low-light object detection algorithm integrating multiple attention mechanisms," *Applied Sciences*, vol. 15, no. 9, p. 5170, 2025.

[24] Z. Han, Z. Yue, and L. Liu, "3L-YOLO: A lightweight low-light object detection algorithm," *Applied Sciences*, vol. 15, no. 1, p. 90, 2024.

[25] A. H. Alsaeedi, A. M. Al-juboori, H. H. R. Al-Mahmood, S. M. Hadi, H. J. Mohammed, M. R. Aziz, M. Aljibawi, and R. R. Nuiaa, "Dynamic clustering strategies boosting deep learning in olive leaf disease diagnosis," *Sustainability*, vol. 15, no. 18, p. 13723, 2023.

[26] A. H. A. Ali Mohsin Al-juboori , Riyadh Rahef Nuiaa, Zaid Abdi Alkareem Alyasseri, Husam Jasim Mohammed, Nor Samsiah Sani, Mohd Isrul Esa, Bashaer Abbuod Musawi, "A Hybrid Cracked Tiers Detection System Based on Adaptive Correlation Features Selection and Deep Belief Neural Networks," *Symmetry*, 2023.

[27] H. Chen, J. Zhang, T. Yu, Y. Zeng, and H. Zeng, "RetinexWT: Retinex-Based Low-Light Enhancement Method Combining Wavelet Transform," *Computers, Materials and Continua*, vol. 86, no. 2, pp. 1-20, 2025/12/09/ 2025, doi: https://doi.org/10.32604/cmc.2025.067041.

[28] Y. Zheng, L. Zhou, K. Jin, S. Jin, Z. Liang, W. Zhao, and W. Zhang, "A comprehensive review of low-light image enhancement methods," *Signal Processing: Image Communication*, vol. 140, p. 117435, 2026/01/01/ 2026, doi: https://doi.org/10.1016/j.image.2025.117435.

[29] B. Mirzaei, H. Nezamabadi-Pour, A. Raoof, and R. Derakhshani, "Small object detection and tracking: a comprehensive review," *Sensors*, vol. 23, no. 15, p. 6887, 2023.