IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 6, December, 2023, pp. 2741-2750
Journal homepage: http://iieta.org/journals/ts

Augmenting Face Detection in Extremely Low-Light CCTV Footage Using the EDCE Enhancement Model

S. Sony Priya*, R. I. Minu

Computing Technologies, SRM Institute of Science and Technology, Kattankulathur 603 203, India

Corresponding Author Email: sp3454@srmist.edu.in

https://doi.org/10.18280/ts.400634

Received: 8 May 2023
Revised: 20 July 2023
Accepted: 12 September 2023
Available online: 30 December 2023

Keywords:
unconstrained video, video enhancement, face detection, zero-reference deep curve estimation, keyframe extraction

ABSTRACT

Face detection constitutes a pivotal task in computer vision, with its utility extending across security and surveillance, biometrics, human-computer interaction, and entertainment. This technology facilitates the automated recognition and location of human faces within images or videos, a feature instrumental for identification, authentication, and tracking. However, the efficacy of face detection algorithms is compromised under low-light conditions prevalent in CCTV videos, due to variations in illumination levels. To address this challenge, this study introduces a video enhancement method, the Enhanced Deep Curve Estimation (EDCE), designed to augment the quality of low-light CCTV footage, thereby improving face detection accuracy. To circumvent the redundancy of frames during face detection from the input video, a key frame extraction method was employed. Subsequently, the Retina Face was utilized to detect faces from the enhanced CCTV video keyframes. The CCTV videos evaluated in this study were sourced from public cameras, and the performance of the EDCE model was assessed against other existing enhancement models. The findings reveal that the EDCE model exhibits superior performance with a Peak Signal-to-Noise Ratio (PSNR) of 21.37 and a Structural Similarity Index Measure (SSIM) of 0.83. Further, the face detection evaluation yielded an Average Precision of 0.847, signifying the effectiveness of our enhancement methodology. This study, thus, underscores the potential of the EDCE model in enhancing the performance of face detection systems under challenging low-light conditions.

1. INTRODUCTION

Face recognition has become increasingly crucial for surveillance applications in various settings, including banks, supermarkets, traffic monitoring systems, criminal identification, crowd monitoring, active device authentication, facial biometrics for payments, and autonomous vehicles. The prerequisite for accurate face identification is high-quality video footage, the quality of which depends on several factors, such as hardware selection, camera positioning and direction, and ambient lighting conditions. Regrettably, lower-resolution CCTV systems often generate pixelated, poorly lit, and noisy images, impeding information extraction and adversely affecting the overall performance of face detection and identification in surveillance systems. Hence, it is imperative to implement low-light enhancement techniques that can offset the performance deficiencies caused by substandard image quality.

Moreover, videos inherently comprise numerous frames, many of which carry redundant information. This redundancy results in multiple detections of the same individual, thereby escalating memory usage, computational resources, and processing time, which represents a significant hurdle in achieving efficient video-based face identification. Furthermore, faces present considerable variations in aspects such as position, lighting, blur, occlusions, and video quality, primarily due to being extracted from unrestricted recordings. Consequently, a robust face detection system must be equipped to manage these divergences and effectively address the challenges intrinsic to face detection.

This paper focuses on videos captured in low-light environments, where face detection presents a formidable challenge. To address this, initial video enhancement is performed. Historically, the Histogram Equalization method [1-5] was employed for image enhancement by adjusting intensity levels. While this method is simple and effective, it is susceptible to noise, which can lead to undesirable outcomes in the enhanced image. The process of equalizing the histogram to adjust intensity levels can sometimes result in the loss of subtle image details. In some instances, Histogram Equalization can introduce unnatural colors or artifacts, which can degrade the overall quality and realism of the enhanced image. Another prevalent enhancement technique is the Retinex theory [6-13], based on the human visual perception system. This enhancement technique considers an image as a composition of reflectance and illumination. However, the Retinex technique is limited in its applicability to real-time applications due to its computational intensity and slow operation for larger images. Additionally, this technique requires multiple images of the same scene captured under

different lighting conditions to calculate color constancy, which can be impractical in certain scenarios.

Deep learning techniques [14-17] have recently surpassed traditional methods with their superior performance and efficiency. However, most of these models are optimized for image enhancement. Video enhancement is a more complex task as it involves processing larger data volumes and dealing with various types of noise and artifacts. In video enhancement, maintaining temporal consistency between frames is critical to prevent flickering. Additionally, these enhancement techniques can be computationally demanding, particularly when dealing with high-resolution videos. As a result, their usage may be limited to real-time applications or devices with constrained processing capabilities. For instance, motion blur can manifest in videos when objects move too fast, causing a loss in detail and sharpness. Furthermore, the limited data available for video enhancement makes it challenging to train deep-learning models.

In this paper, we introduce a novel model, the Enhanced Deep Curve Estimation (EDCE), which builds upon the DCE-Net [18] architecture. Like DCE-Net, EDCE is a lightweight network that does not require paired or unpaired data for training. We propose a cubic enhancement function within the EDCE model, establishing a stronger non-linear association between the input and enhanced image. This function effectively preserves critical features and enhances visual details, thereby improving overall image quality.

Upon applying the EDCE model for video enhancement, we proceed with Key-Frame extraction. This module enables us to identify and extract the key frames from the video, eliminating redundant frames and focusing exclusively on the essential ones. By processing only these key frames, we can achieve higher accuracy in face detection while simultaneously reducing computational complexity and storage needs. For the Key-Frame extraction process, we utilize a CNN-based model.

To identify faces within the enhanced video footage, we deploy the Retina Face model [19], a state-of-the-art face detection model. This final step aims to demonstrate the efficacy of our proposed image enhancement model in enhancing the accuracy of face detection. By integrating the EDCE enhancement model, Key-Frame extraction, and Retina Face, we aim to enhance video quality and optimize face detection performance in low-light environments.

The primary contributions of this paper are as follows:

- Enhanced low-light CCTV videos using the EDCE (Enhanced Deep Curve Estimation) model.
- Proposed a CNN-based Key Frame extraction module to avoid redundant frames.
- Demonstrated face detection results using both existing and proposed video enhancement methods.

The remaining sections of the paper are organized as follows. Section 2 provides an overview of related works in the field of low-light video enhancement and Keyframe extraction. Section 3 outlines the employed methodology, which includes details about the EDCE model and the CNN-based Keyframe extraction module. Section 4 describes the experimentaldetails, including the dataset used and the evaluation metrics. Section 5 presents the results obtained from the experiments, comparing the performance of the proposed method to existing approaches. Finally, Section 6 draws conclusions based on the study's findings and discusses potential areas for future research in this domain.

2. RELATED WORKS

Deep learning techniques have recently surpassed traditional methods such as Histogram Equalization, Retinex Theory-based models, and Gamma correction [20, 21]. Many of these conventional methods primarily focus on enhancing the image's brightness, but often neglect the issue of noise amplification. As a result, increasing the brightness often simultaneously increases the noise, leading to a degradation in image quality. Some methods, like Retinex-based models, perform enhancement and denoising separately. However, if enhancement precedes denoising, it can amplify noise, and if denoising precedes enhancement, it can result in image blurring.

To overcome the problems of blurring and noise, Lv et al. [22] proposed an attention-guided low light enhancement model that simultaneously enhances brightness and removes noise. However, this model struggles to capture facial details within highly dark regions of the image and can create blocking artifacts due to heavy compression. Che Aminudin and Suandi [23] proposed a deep learning model that utilizes CNN and an autoencoder model to tackle issues of low illumination, low resolution, and noise. Anitha and Kumar [24] focused on improving image resolution and illumination while reducing noise using a GAN model. However, this model is particularly effective for images taken indoors.

Retinex DIP (Deep Image Prior) model, proposed by Zhao et al. [25], integrates Retinex theory and neural networks. This model enhances the low-light input image by obtaining illumination information through Retinex decomposition. Notably, this model only requires a single low-light image for training and does not rely on any external datasets. Lamba et al. [26] introduced LLPackNet, which enhances the image while reducing memory utilization and model parameters. Although this model reduces processing time, it can introduce blurring for larger down-sampled images.

Li et al. [27] proposed a lightweight Zero DCE (Deep curve Estimation) network that eliminates the need for paired and unpaired data. Despite its merits, like any deep learning-based method, this model requires a large training dataset and careful parameter tuning to achieve optimal results. Furthermore, the enhancement curve of this image enhancement model is linear, which may not capture complex non-linear relationships. We propose a quadratic enhancement curve to address this limitation, where the output is computed using a quadratic function of the input image.

In the field of CCTV monitoring, keyframe extraction is a valuable tool that selects a representative subset of frames from a video sequence. This can greatly assist in identifying and tracking faces in video footage, facilitating efficient analysis of large volumes of CCTV footage. Muhammad et al. [28] extracted keyframes using a combination of Mobile Net architecture, memorability and entropy scores, and color histograms. However, the processing rate of 18 fps may not suffice for real-time video surveillance applications. Basavarajaiah, and Sharma [29] performed keyframe extraction by selecting keyframes specifically when a significant scene change was detected within the video. Yuan et al. [30] proposed a Global Motion Statistics-based Scheme (KEGMS) that extracts keyframes based on global motion and events. Yasmin et al. [31] proposed a novel agglomerative clustering algorithm for extracting informative frames based on key moments. However, this model may require high computation time for larger input video files. In this study, we

adopt a technique inspired by Basavarajaiah and Sharma [29] due to its efficiency and speed, incorporating a pre-trained model to extract features from video frames and a straightforward frame selection algorithm for video summarization.

3. METHODOLOGY

The primary focus of this paper is to enhance the quality of CCTV videos captured in low-light conditions. The quality of these videos is influenced by assorted aspects such as the camera's quality, position, lighting conditions, and the distance between the camera and the human face. Therefore, a novel video enhancement is performed first to enhance the dark videos. Then, from this enhanced video, the keyframes are selected to avoid processing repeated frames. Finally, the face is detected by using the pre-trained model RetinaFace.

3.1 Video enhancement

Inspired by the work of Guo et al. [18], we proposed EDCE (Enhanced Deep Curve Estimation) to enhance the low light input CCTV video. Like the zero DCE model, the EDCE model also contains seven convolutional layers, each featuring 32 filters with a dimension of 3×3 and a stride of 1With the exception of the last layer, the ReLU activation function is applied to all the other layers. The Tanh activation function is specifically used in the last convolutional layer. Furthermore, the model takes an image as input and generates higher-order curves to facilitate image enhancement. In the paper [18], the authors used the below formula for enhancement:

$$LEC((I(i);\gamma)=I(i)+\gamma I(i)(1-I(i)) \quad (1)$$

Here I(i) denotes the pixel in the input image, 1-I(i) represents its complement, and γ is the learned parameter that controls the enhancement amount applied to the input low-light image learned during training. A higher value of γ results in a stronger enhancement effect, while a lower value of γ results in a more subtle effect. However, this model fails to enhance the extremely low-light images. Hence, we changed Eq. (1). to find the enhancement curve to capture more complex variations in the input image. Eq. (2). is the modified low light enhancement curve.

$$LEC((I(i):\gamma) = I(i) + \gamma I(i)^2(1 - I(i))^2 \quad (2)$$

To apply the above equation for video, first, we convert the video into frames then it is passed to the enhanced DCE model to perform an enhancement. Here the enhancement curve is repeatedly applied to the enhanced until we get the expected enhanced frame. For that, the difference between the enhanced frame from the current iteration and the previous iteration is calculated. Before this step, we have to fix the threshold value in the initial stage. Here the specified threshold value is 0.3. If the difference is below the threshold, the enhancement process is stopped, and the final enhanced image is returned as the final result of the processing. On the other hand, if the difference is above the threshold, the enhancement process continues with another iteration. Figure 1 shows the complete architecture of video enhancement.

3.1.1 Loss functions

To improve the quality of the enhanced frames, loss functions are utilized. In this paper, we utilized four different loss functions, including the spatial consistent loss function, color constancy loss function, exposure control loss, and total variation loss.

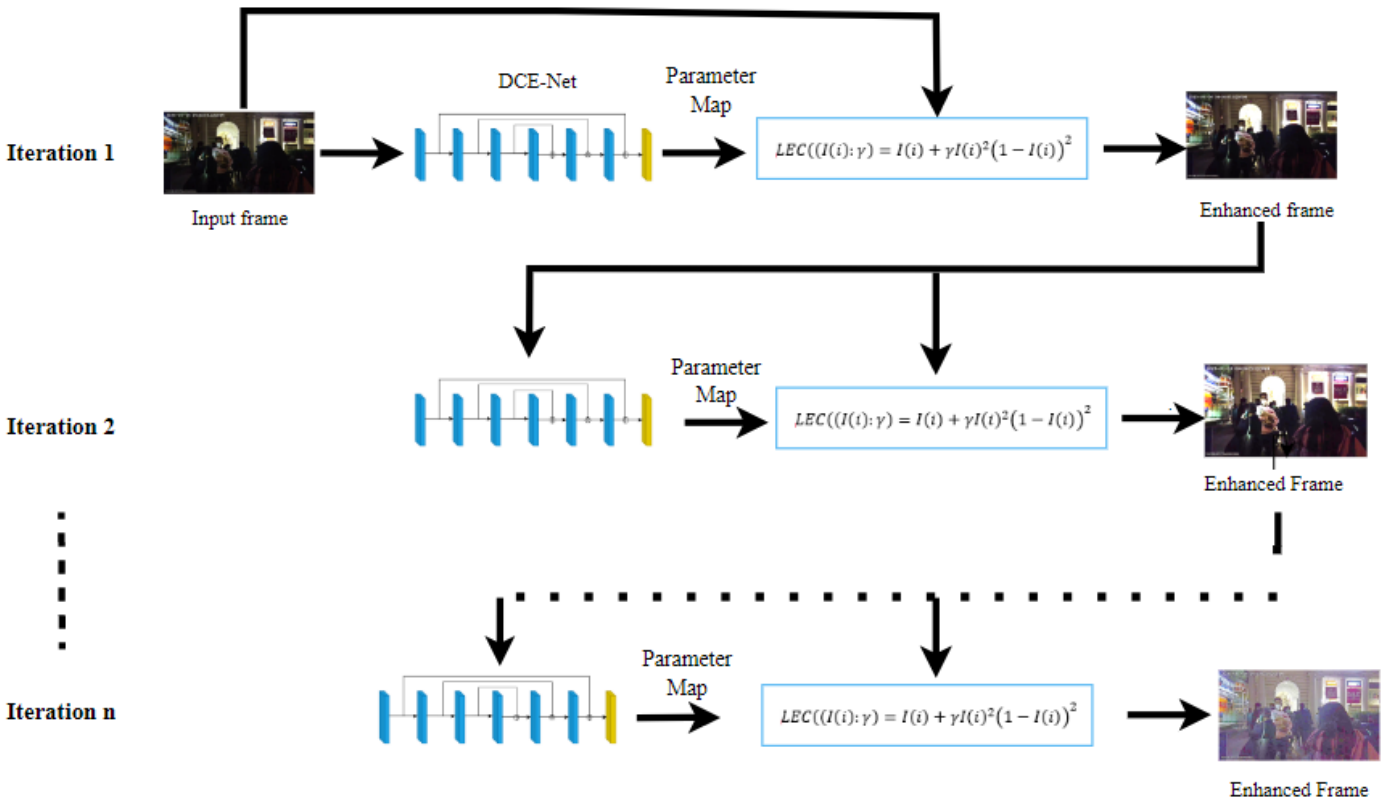


Figure 1. EDCE architecture

Spatial consistent loss function (L_{sc})

The non-reference Spatial consistency loss function aims to maintain spatial consistency between the input and enhanced images by minimizing the differences in their spatial structures. This loss function can be mathematically represented as:

$$L_{sc} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j\in R(i)}(|(E_i - E_j)| - |(I_i - I_j)|)^2 \quad (3)$$

Here, *N* refers to the number of smaller regions partitioned from the image, while *R* represents the four conterminous regions incorporated at *i.* The variables *E* and *I* denote the output image after enhancement and the original image before enhancement.

Color constancy loss function (L_{cc})

Non-reference color constancy loss is a loss function commonly employed in image enhancement algorithms to assess the color constancy of the enhanced image. Unlike reference-based methods, non-reference methods do not require a reference image to evaluate color constancy. Instead, they rely on statistical measures such as the Mean and Standard-deviation of pixel values to estimate the degree of color constancy in an image.

$$L_{col} = \sum_{\forall(p,q)\in(R,G),(R,B),(G,B)}\left(E_p - E_q\right)^2 \quad (4)$$

where, E_p and E_q are the pixel values of the enhanced image at the corresponding positions in the color channels p and q.

Exposure control loss (L_E)

This loss function aims to balance the exposure level across different partitions of the image by computing the difference between the intensity values of these partitions and the Well-exposedness Measure (WEM). Here, value of *E* is measured by using the Gauss curve [32]. The loss function is defined as:

$$L_E = \frac{1}{N}\sum_{k=1}^{N}|E_k - WEM| \quad (5)$$

Total variation loss (L_{tv})

Total Variation (TV) loss is used to reduce the amount of noise and artifacts in the enhanced image by minimizing the total variation of the image. Total variation measures the amount of variation in the intensity of adjacent pixels in the image. The TV loss encourages the enhanced image to have smooth and continuous regions by penalizing abrupt changes in the intensity of adjacent pixels.

$$L_{tv}(e) = \sqrt{(e_{i+1,j} - x_{i,j})^2 + (e_{i,j+1} - x_{i,j})^2} \quad (6)$$

where, *e* denotes enhanced image, and *i* and *j* represent the pixel coordinates in the image. Finally, the total loss function is given by:

$$L_{TL} = L_{sc} + L_E + w_{cc}L_{cc} + w_{tv}L_{tv} \quad (7)$$

The weight factors w_{cc}, w_{tv} controls the trade-off between reducing noise and preserving fine details in the enhanced image.

3.2 Key frame extraction

In this research paper, we incorporated a keyframe extraction module to address the issue of processing repeated frames. As videos often contain redundant frames that do not contribute substantially to the overall content, selecting only frames that exhibit changes becomes crucial for efficient analysis. Here, we incorporated a keyframe extraction module based on a previous study [29] to optimize processing time and memory usage. For feature extraction, a pre-trained VGG16 model [33] was employed to extract features from the video. The output of the final convolutional layer provided the feature representation for each frame. Subsequently, k-means clustering was applied to group the feature vectors of the frames. Keyframes were identified by selecting frames that were closest to the centroids of each cluster. These keyframes were then saved as individual images.

Initially, the number of clusters was set to 5, and the optimal number of clusters was determined using the elbow method. Algorithm 1 and Figure 2 provide a comprehensive illustration of the keyframe extraction process, detailing the essential steps involved. The integration of this keyframe extraction module aimed to improve the efficiency of video analysis by focusing on frames that capture significant changes. This approach resulted in a more concise and informative video summary, reducing the processing of redundant frames.

Algorithm

Input: Video (V), Pre-trained CNN model (M), Number of keyframes to extract (K)

Output: Set of K keyframes (K_F)

Step 1: Extract features from the video:

Decompose video into frames {f_1, f_2, f_3,……,f_n}

for each frame f in V:

Pass f into the Pretrained CNN model M to obtain feature map $\mathcal{F}$

Flatten $\mathcal{F}$ to a 1D feature vector fv.

Store fv in a feature array F[].

Step 2: Determine the optimal number of clusters using the elbow method:

Initialize an empty list of distortion values D.

for k in range (1, max(k+1)):

Apply K-means clustering to F[] with k clusters, obtaining cluster labels L.

Compute the distortion value for this clustering using the sum of squared distances between every point and its assigned centroid.

Append the distortion value to D.

Plot D as a function of k and identify the elbow point as the optimal number of clusters.

Step 3: Apply K-means clustering to F [] with K clusters, obtaining cluster labels L.

Select the keyframes from each cluster:

For each cluster c in L:

Compute the distance between each feature vector in c and the cluster centroid.

Sort the feature vectors in c by their distance to the centroid.

Select the top-ranked feature vectors as the keyframes for this cluster.

Map each keyframe's feature vector back to its corresponding video frame and add it to the set of keyframes KF.

Return the set of K keyframes KF.

3.3 Detection

After extracting the keyframes, faces are detected from it by using the pre-trained face detection model RetinaFace [19]. Retina Face is a face detection algorithm that can work in real-world scenarios where faces vary in scale, pose, and occlusion.

It is a single-stage detector that uses a dense anchor box design and a multi-task loss function to simultaneously predict the bounding boxes of each detected face. The algorithm is based on a modified version of the Single Shot Detector (SSD) [34] architecture, which uses a single convolutional network to predict object detections. RetinaFace incorporates a novel scale-aware training strategy that helps the algorithm better handle faces of different sizes.

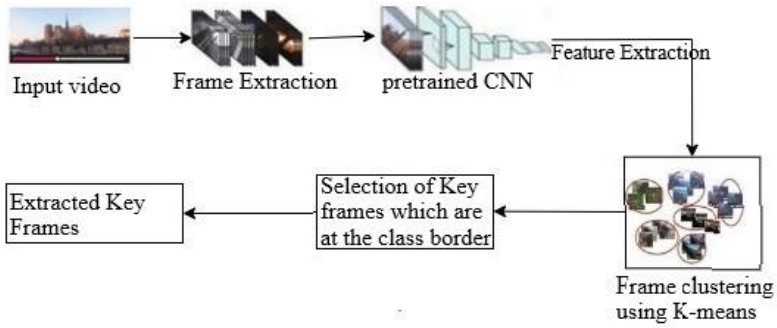


Figure 2. Key frame extraction

4. EXPERIMENTAL DETAILS

The dataset we collected consists of a total of 219 videos, with a duration of 30s. The resolution of videos is 1920×1080. Due to the difficulty in obtaining real nighttime CCTV videos without light and with several people, we simulated low-light conditions by darkening videos captured during the period of 6 pm to 10 pm using the OpenCV library.

For video enhancement, we implemented the EDCE model using PyTorch on Google Colab Pro. The model was trained for 50 epochs with the Adam optimizer and a learning rate of 0.0001. The goal was to improve the quality of the low-light videos and enhance important features while preserving visual details. After enhancing the videos, we utilized a keyframe extraction module based on a pre-trained VGG16 model. This module selected keyframes that represented significant changes in the video content. These keyframes were identified by applying k-means clustering to the feature vectors of the frames and selecting frames closest to the centroids of each cluster. Finally, we employed a Retina Face detection model to detect faces in the enhanced keyframes. This model was applied to identify and locate faces accurately within the low-light video footage. The combination of video enhancement, keyframe extraction, and face detection aimed to improve the overall quality of the video to detect faces.

In this study, we computed PSNR (Peak Signal to Noise Ratio) and SSIM (Structural Similarity Measure) [35] for evaluation purpose. PSNR is a widely used performance metric to evaluate the quality of enhanced images or videos. It measures the dissimilarity between the original and enhanced frames by computing the ratio between the maximum possible pixel value (peak signal) and the Mean Squared Error (MSE) of the two images. A higher PSNR value indicates that the enhanced video is more similar to the original. The formula to calculate PSNR value is given by:

$$PSNR = 20\log_{10}\left(\frac{MAX_p}{\sqrt{MSE}}\right) \tag{8}$$

$$MSE = \frac{1}{h,w}\sum_0^{h-1}\sum_0^{w-1}(O(x,y) - E(x,y))^2 \tag{9}$$

In the PSNR equation, MAXp represents the maximum possible pixel value of the image (usually 255 for 8-bit images), while h and w represent the height and width of the image. O(x,y) represents the pixel value of the original image at position (x,y), while E(x,y) represents the pixel value of the enhanced image at the same position.

On the other hand, SSIM considers not only the pixel values but also the structural information of the video. It compares the luminance, contrast, and structure of the original and enhanced video frames and calculates their similarity. The SSIM value ranges between 0 and 1, with 1 indicating perfect similarity. SSIM is known to correlate better with subjective visual quality than PSNR. The formula for calculating SSIM is:

$$\mathrm{SSIM(o,e)} = \mathrm{L(o,e)} * \mathrm{C(o,e)} * \mathrm{S(o,e)} \tag{10}$$

where, *o* and *e* are the original and enhanced images, respectively. *L(o,e), C(o,e),* and *S(o,e)* are the luminance, contrast, and structure similarity measures, respectively, and are defined as:

$$L(\mathrm{o,e}) = \frac{(2 * \mu_o * \mu_e + c_1)}{(\mu_o{}^2 + \mu_e{}^2 + c_1)} \tag{11}$$

$$C(\mathrm{o,e}) = \frac{(2*\sigma_o\,\sigma_e + c_2)}{(\sigma_o{}^2 + \sigma_e{}^2 + c_2)} \tag{12}$$

$$S(\mathrm{o,e}) = \frac{\sigma_{oe} + c_3}{\sigma_o\,\sigma_e + c_3} \tag{13}$$

where, μ_o and μ_e represent the mean pixel values of *o* and *e*, respectively, similarly, σ_o and σ_e are the standard deviations of the pixel values of o and e, respectively, while σ_{oe} is the covariance of the pixel values of *o* and *e*, and c_1, c_2, and c_3 are small constants to prevent division by zero.

5. RESULTS

The proposed model was compared with existing enhancement methods ZeroDCE [18], Enlighten GAN [36], LIME [11], Semantic-Guided zero-shot learning [37] to evaluate its performance. Figure 3 shows the sample frames of the input video. The hyperparameter settings are explained in section 4. Figure 4 shows the training and validation loss curves using the Enhanced DCE Video Enhancement model. For comparison, we used already published codes available on

GitHub.

The results of various enhancement models are visually shown in Figure 5. From the resulting images, we can understand the EDCE video enhancement model provides better result as compared with other methods. Table 1 and Figure 6 presents the performance evaluation of the models using the metrics PSNR and SSIM. From this, we can understand our model is better than other models in extremely dark video enhancement.

Figure 3. Sample low-light CCTV video frames

After video enhancement, key frame extraction is performed. Figure 7 represents the output of keyframe extraction. By doing this, we can avoid repeated frames to detect faces. Here, we applied Retina Face to detect faces. Figure 8 illustrates the results of face detection in the enhanced video frames mentioned above. From the results, we can understand EDCE method outperforms the state-of-art enhancement methods. The proposed model not only increases the brightness of the image but also preserves its color, and it also provides a higher face detection rate. Figure 9 shows the training and validation accuracy of the Retina face detection model where the input frame is enhanced by using the proposed EDCE model. Figure 10 shows the Precision-Recall curve for face detection using different enhancement model images. However, a drawback of this model is the occurrence of flickering in the enhanced video. Flickering occurs when there are abrupt and inconsistent changes between consecutive frames after enhancement, causing visual disturbances in the video. To avoid flickering problems in video enhancement, you need to ensure temporal consistency across frames.

Table 1. Quantitative evaluation of different enhancement methods

Model	PSNR	SSIM
ZeroDCE [2]	16.57	0.59
MBLLEN [38]	20.56	0.71
Enlighten GAN [17]	16.21	0.59
LIME [18]	16.17	0.57
Semantic-Guided-Low-Light-Image-Enhancement [19]	16.60	0.613
EDCE	21.37	0.83

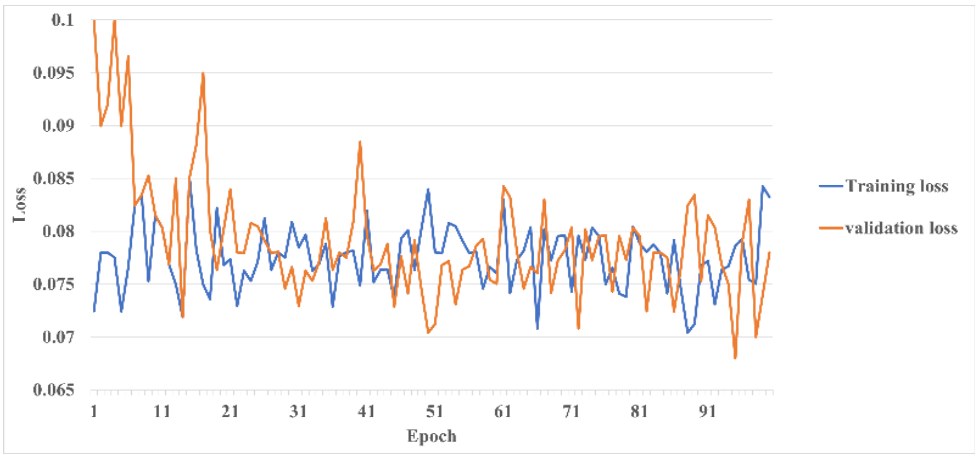


Figure 4. Training and validation loss using the EDCE model

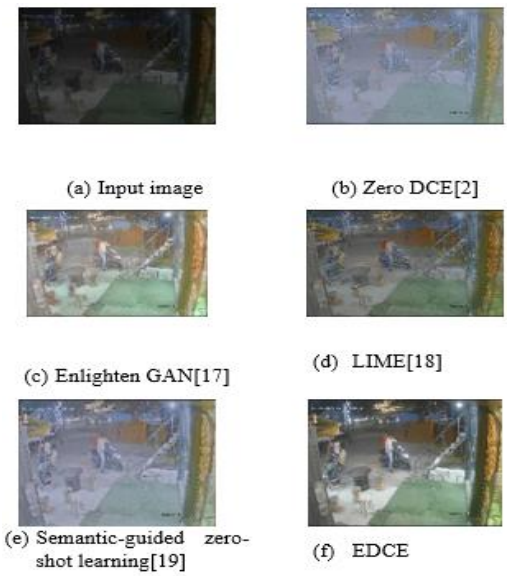
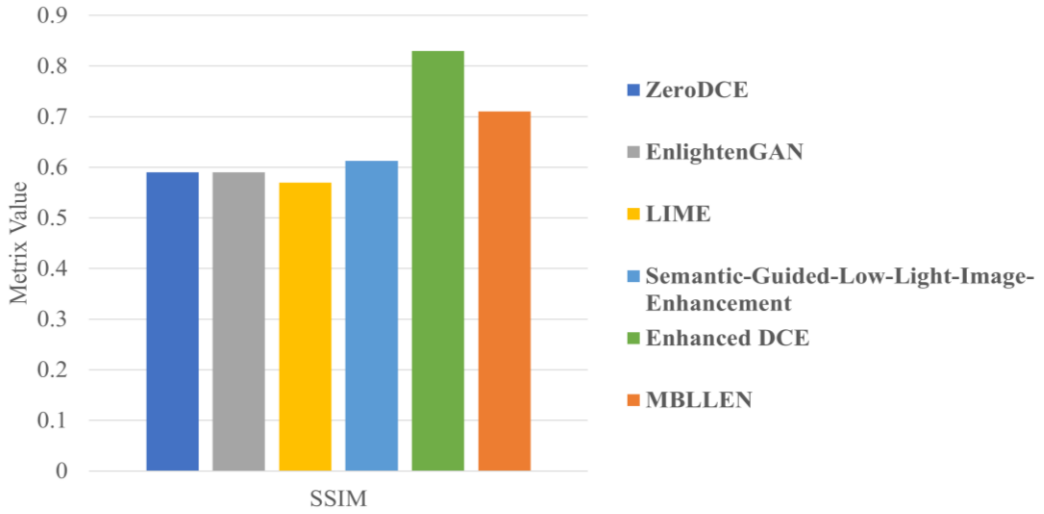
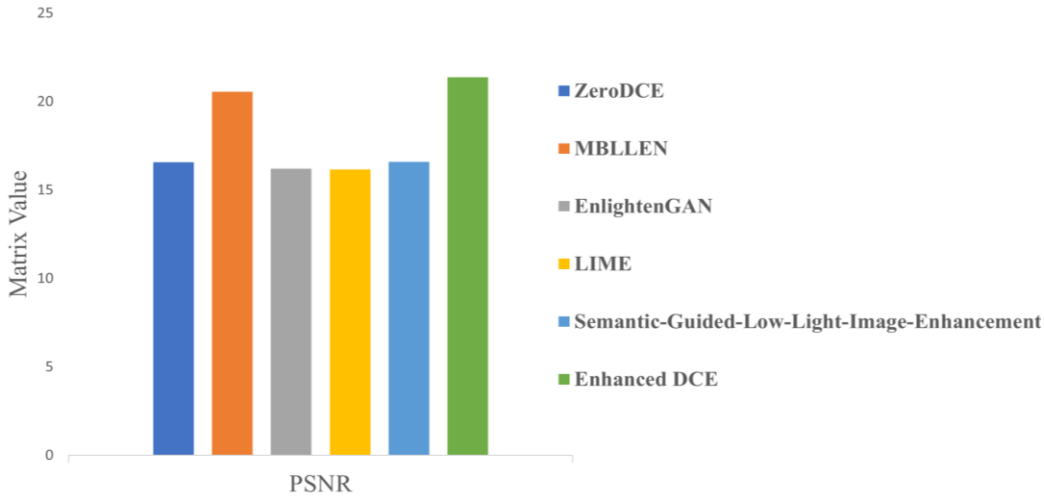


Figure 5. Results of different enhancement model



(a) SSIM comparison



(b) PSNR comparison

Figure 6. Graphical representation of performance evaluation

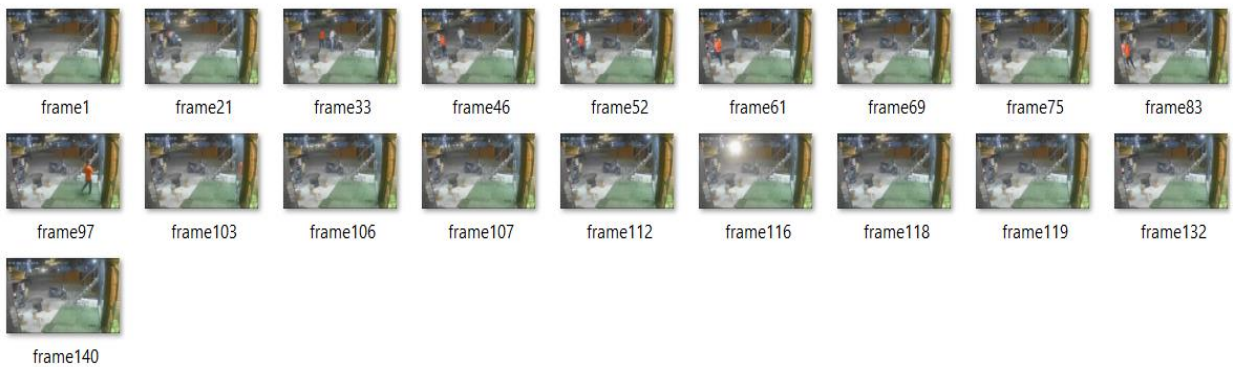
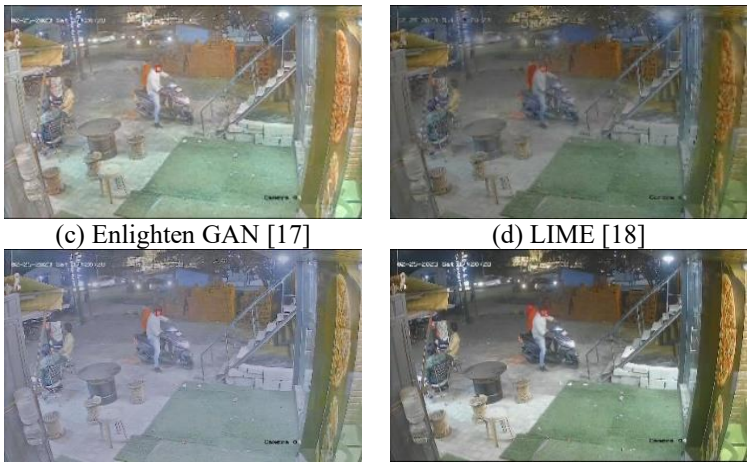


Figure 7. Sample output for keyframe extraction

(a) Input image (b) Zero DCE [2]



(c) Enlighten GAN [17] (d) LIME [18]

(e) Semantic-guided zero-shot learning [19] (f) EDCE

Figure 8. Results of face detection in various enhancement results

Figure 9. Training and validation accuracy of Retina Face detection model

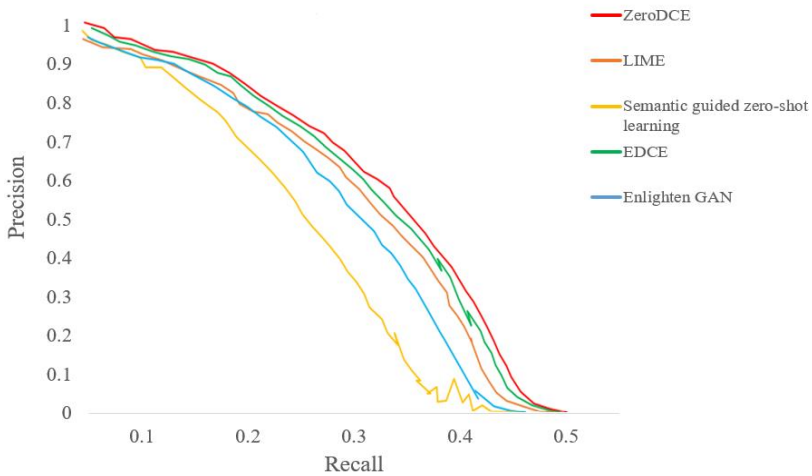


Figure 10. PR curve for face detection

6. CONCLUSION

In this paper, we proposed an EDCE model based on the Zero-DCE model to enhance low-light CCTV videos. Here, we used a cubic enhancement curve to enhance the video. From the visual results, we can observe that this model not only enhances the curve but also preserves the color details of the image. From the quantitative analysis, we can find that the PSNR and SSIM values are also high. After the enhancement, key frame extraction is performed to avoid repeated frames.

Then face detection is performed by using the RetinaFace detection method. With this enhancement, the face detection rate is also increased due to the higher image details. The enhanced video footage can greatly benefit security personnel, law enforcement, and other stakeholders involved in video monitoring, as it provides clearer and more informative visual data for analysis. This, in turn, can lead to quicker response times and more effective decision-making in critical situations. To further enhance the model's practical impact and address the flickering issue observed in the current implementation, several potential techniques and approaches can be explored. For example, temporal smoothing techniques can be employed to reduce flickering by considering the information from adjacent frames and creating smoother transitions between them.

REFERENCES

[1] Pizer, S.M., Amburn, E.P., Austin, J.D., Cromartie, R., Geselowitz, A., Greer, T., Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. Computer Vision, Graphics, and Image Processing, 39(3): 355-368. https://doi.org/10.1016/S0734-189X(87)80186-X

[2] Yadav, G., Maheshwari, S., Agarwal, A. (2014). Contrast limited adaptive histogram equalization based enhancement for real time video system. In 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Delhi, India, pp. 2392-2397. https://doi.org/10.1109/ICACCI.2014.6968381

[3] Kim, Y.T. (1997). Contrast enhancement using brightness preserving bi-histogram equalization. IEEE transactions on Consumer Electronics, 43(1): 1-8. https://doi.org/10.1109/30.580378

[4] Pisano, E.D., Zong, S., Hemminger, B.M., DeLuca, M., Johnston, R.E., Muller, K., Pizer, S.M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. Journal of Digital Imaging, 11: 193-200. https://doi.org/10.1007/BF03178082

[5] Reza, A.M. (2004). Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology, 38: 35-44. https://doi.org/10.1023/B:VLSI.0000028532.53893.82

[6] Land, E.H. (1977). The retinex theory of color vision. Scientific American, 237(6): 108-129.

[7] Jobson, D.J., Rahman, Z.U., Woodell, G.A. (1997). Properties and performance of a center/surround retinex. IEEE Transactions on Image Processing, 6(3): 451-462. https://doi.org/10.1109/83.557356

[8] Jobson, D.J., Rahman, Z.U., Woodell, G.A. (1997). A multiscale retinex for bridging the gap between color images and the human observation of scenes. IEEE Transactions on Image Processing, 6(7): 965-976. https://doi.org/10.1109/83.597272

[9] Wang, S., Zheng, J., Hu, H.M., Li, B. (2013). Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing, 22(9): 3538-3548. https://doi.org/10.1109/TIP.2013.2261309

[10] Fu, X., Zeng, D., Huang, Y., Liao, Y., Ding, X., Paisley, J. (2016). A fusion-based enhancing method for weakly illuminated images. Signal Processing, 129: 82-96. https://doi.org/10.1016/j.sigpro.2016.05.031

[11] Guo, X., Li, Y., Ling, H. (2016). LIME: Low-light image enhancement via illumination map estimation. IEEE Transactions on Image Processing, 26(2): 982-993. https://doi.org/10.1109/TIP.2016.2639450

[12] Kimmel, R., Elad, M., Shaked, D., Keshet, R., Sobel, I. (2003). A variational framework for retinex. International Journal of Computer Vision, 52: 7-23. https://doi.org/10.1023/A:1022314423998

[13] Fu, X., Zeng, D., Huang, Y., Ding, X., Zhang, X.P. (2013). A variational framework for single low light image enhancement using bright channel prior. In 2013 IEEE Global Conference on Signal and Information Processing, Austin, TX, USA, pp. 1085-1088. https://doi.org/10.1109/GlobalSIP.2013.6737082

[14] Lore, K.G., Akintayo, A., Sarkar, S. (2017). LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition, 61: 650-662. https://doi.org/10.1016/j.patcog.2016.06.008

[15] Shen, L., Yue, Z., Feng, F., Chen, Q., Liu, S., Ma, J. (2017). Msr-net: Low-light image enhancement using deep convolutional network. arXiv preprint arXiv:1711.02488. https://doi.org/10.48550/arXiv.1711.02488

[16] Ren, W., Liu, S., Ma, L., Xu, Q., Xu, X., Cao, X., Yang, M.H. (2019). Low-light image enhancement via a deep hybrid network. IEEE Transactions on Image Processing, 28(9): 4364-4375. https://doi.org/10.1109/TIP.2019.2910412

[17] Guo, Y., Ke, X., Ma, J., Zhang, J. (2019). A pipeline neural network for low-light image enhancement. IEEE Access, 7: 13737-13744. https://doi.org/10.1109/ACCESS.2019.2891957

[18] Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R. (2020). Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, pp. 1780-1789. https://doi.org/10.1109/CVPR42600.2020.00185

[19] Deng, J., Guo, J., Ververas, E., Kotsia, I., Zafeiriou, S. (2020). Retinaface: Single-shot multi-level face localisation in the wild. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Seattle, WA, USA, pp. 5203-5212. https://doi.org/10.1109/CVPR42600.2020.00525

[20] Zheng, L., Shi, H., Gu, M. (2017). Infrared traffic image enhancement algorithm based on dark channel prior and gamma correction. Modern Physics Letters B, 31(19-21): 1740044. https://doi.org/10.1142/S0217984917400449

[21] Chang, Y., Jung, C., Ke, P., Song, H., Hwang, J. (2018). Automatic contrast-limited adaptive histogram equalization with dual gamma correction. IEEE Access, 6: 11782-11792. https://doi.org/10.1109/ACCESS.2018.2797872

[22] Lv, F., Li, Y., Lu, F. (2021). Attention guided low-light image enhancement with a large scale low-light simulation dataset. International Journal of Computer Vision, 129(7): 2175-2193. https://doi.org/10.1007/s11263-021-01466-8

[23] Che Aminudin, M.F., Suandi, S.A. (2022). Video

surveillance image enhancement via a convolutional neural network and stacked denoising autoencoder. Neural Computing and Applications, 34: 3079-3095. https://doi.org/10.1007/s00521-021-06551-0

[24] Anitha, C., Kumar, R.M.S. (2022). GEVE: A generative adversarial network for extremely dark image/video enhancement. Pattern Recognition Letters, 155: 159-164. https://doi.org/10.1016/j.patrec.2021.10.030

[25] Zhao, Z., Xiong, B., Wang, L., Ou, Q., Yu, L., Kuang, F. (2021). RetinexDIP: A unified deep framework for low-light image enhancement. IEEE Transactions on Circuits and Systems for Video Technology, 32(3): 1076-1088. https://doi.org/10.1109/TCSVT.2021.3073371

[26] Lamba, M., Balaji, A., Mitra, K. (2020). Towards fast and light-weight restoration of dark images. arXiv preprint arXiv:2011.14133. https://doi.org/10.48550/arXiv.2011.14133

[27] Li, C., Guo, C., Loy, C.C. (2021). Learning to enhance low-light image via zero-reference deep curve estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(8): 4225-4238. 10.1109/TPAMI.2021.3063604

[28] Muhammad, K., Hussain, T., Baik, S.W. (2020). Efficient CNN based summarization of surveillance videos for resource-constrained devices. Pattern Recognition Letters, 130: 370-375. https://doi.org/10.1016/j.patrec.2018.08.003

[29] Basavarajaiah, M., Sharma, P. (2021). GVSUM: generic video summarization using deep visual features. Multimedia Tools and Applications, 80: 14459-14476. https://doi.org/10.1007/s11042-020-10460-0

[30] Yuan, Y., Lu, Z., Yang, Z., Jian, M., Wu, L., Li, Z., Liu, X. (2022). Key frame extraction based on global motion statistics for team-sport videos. Multimedia Systems, 28(2): 387-401. https://doi.org/10.1007/s00530-021-00777-7

[31] Yasmin, G., Chowdhury, S., Nayak, J., Das, P., Das, A.K. (2023). Key moment extraction for designing an agglomerative clustering algorithm-based video summarization framework. Neural Computing and Applications, 35(7): 4881-4902. https://doi.org/10.1007/s00521-021-06132-1

[32] Mertens, T., Kautz, J., Van Reeth, F. (2007). Exposure fusion. In 15th Pacific Conference on Computer Graphics and Applications (PG'07), Maui, HI, USA, pp. 382-390. https://doi.org/10.1109/PG.2007.17

[33] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

[34] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. (2016). SSD: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, pp. 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[35] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600-612. https://doi.org/10.1109/TIP.2003.819861

[36] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Wang, Z. (2021). Enlightengan: Deep light enhancement without paired supervision. IEEE Transactions on Image Processing, 30: 2340-2349. https://doi.org/10.1109/TIP.2021.3051462

[37] Zheng, S., Gupta, G. (2022). Semantic-guided zero-shot learning for low-light image/video enhancement. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, pp. 581-590. https://doi.org/10.1109/WACVW54805.2022.00064

[38] Lv, F., Lu, F., Wu, J., Lim, C. (2018). MBLLEN: Low-Light Image/Video Enhancement Using CNNs. In BMVC, 220(1): 1-13.