EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Enhancement of Criminal Facial Image Using Multistage Progressive V-Net for Facial Recognition by Pixel Restoration

S.S. Beulah Benslet[1,*] and P. Parameswari[2]

[1]Research Scholar, Department of Computer Science, Karuppannan Mariappan College, Muthoor, Chettiyarpalayam
[2]Principal, Palanisamy College of Arts, Perundurai-Erode Road, Tamil Nadu 638052, India

Abstract

INTRODUCTION: Criminal activity is expanding exponentially in modern society, which leads towards a great concern about security issues. Facial recognition technology (FRT) is a powerful computer-based system that increasingly being used for recognize and match faces to solve crimes and investigations.
OBJECTIVES: Due to poor image clarity and noisy pixels, the detection of criminal faces tends to be inaccurate. Hence, image enhancement techniques are required to recognize criminals with better accuracy. In the proposed model, a multistage progressive V-net based image quality enhancing technique is employed to improve accuracy.
METHODS: The Convolutional Neural Network (CNN) for restoring images called MPRV-Net has three stages for a difficult balance between spatial data and highly contextualized information for image restoration tasks while recovering images.
RESULTS: For image restoration tasks, including denoising, deblurring, and deraining, MPRV-Net has provided considerable performance benefits on a number of datasets. The suggested network is significant as it eliminates all three types of deviations using a single architecture. The proposed model's performance is tested using performance metrics such as accuracy, precision, recall, and specificity, obtaining 94%, 96%, 93%, and 95%.
CONCLUSION: Thus, the proposed Multistage Progressive V-Net model for effectively improves the criminal Facial image for detecting criminals in public places with greater accuracy.

Keywords: criminal, facial recognition technology, image restoration, image enhancement, MPRV-Net

Received on 26 September 2023, accepted on 10 December 2023, published on 18 January 2024

doi: 10.4108/eetsis.3980

1. Introduction

Criminal identification is the most significant but challenging and time-consuming task for the police. It will be challenging in crowded environments, such as cities or public places and in some cases, manual identification techniques provide the possibility to acquire more details about offenders. Still, it's time-consuming and involves the risk of missing criminals. Public supervision is not necessary for an automated identification system (AIS) [1]. This will help the police in tracking down and apprehending criminals in public areas. Relevant face detection and recognition algorithms are included in automated identification systems [2]. Face recognition for criminal identification uses a distinct biometric method. This approach identifies and verifies a person's identity based on video or image frames, including their face. Face recognition preprocessing is integrated systems that enhance an input face image to improve the quality of the image by increasing the visibility of facial features and improving prediction performance. The

*Corresponding author. Email: beulahabraham15@gmail.com

efficiency of face recognition systems is improved by pre-processing [3]. Additionally, the preprocessing stage improves distorted images and gathers regions of interest in an image for further feature extraction.

The pre-processing step of image enhancement is necessary for enhancing the performance of face recognition systems. Face image enhancers are more effective approaches used before the recognition process and are commonly included in the design of most face recognition systems [4]. Face recognition systems employ a wide range of image enhancement techniques. Each of these approaches has advantages as well as disadvantages. Some recent methods modify an input image resulting in a more precise or detailed output image [5]. However, in uncontrolled contexts, the quality of images in face recognition systems may be significantly decreased for numerous reasons, including lighting conditions, i.e., in dark or overly bright environments. Restoring a clear image from an imperfect one is known as image restoration [6]. CNN-based methods outperform other approaches primarily owing to model development. Among the computational modules and functional units designed for image recovery include encoder-decoders [11], generative models [12], dilated convolutions [8, 9], dense connections [10], and recursive residual learning [7].

Recently, there haven't been many attempts to use the multi-stage method for image deraining and deblurring [13]. Investigate these methods to discover the architectural constraints limiting their efficacy [14]. In order to begin, current multi-stage techniques either rely on a single-scale pipeline [16], which produces spatially accurate but semantically less reliable outputs, or an encoder-decoder architecture [15], which is reliable in preserving broad contextual information but unreliable in preserving spatial image details. However, the conventional approach is insufficient to differentiate between criminal and non-criminal behaviour. The majority of the accessible models for low-level visual issues are developed in a single stage. In high-level performance, however, it has been proved that multi-stage networks outperform single-stage analogues. To increase accuracy, the proposed model has introduced multistage progressive v-net based image restoration. The following is presented as the proposed model's contribution:

- Multistage progressive V net had been proposed for pixel restoration, effectively enhancing criminal images.
- The collected dataset for criminal recognition is gathered from public websites and split into criminal and non-criminal categories for training classifiers.
- Vnet architecture is used in multistage progressive networks for extracting the pixel level feature based on encoding and decoding.
- The final stage includes the original resolution sub network (ORSNet) to achieve spatially relevant outcomes for a pixel-to-pixel correspondence between the input and output images.
- An efficient Supervised Attention Module (SAM) that uses the restored image at every level to enhance incoming features before propagating them further.

The remaining portion of the manuscript is sectioned as follows: Section 2 denotes some related work related to criminal prediction. Section 3 presents the proposed methodology along to predict criminals. Section 4 presents the result and discussion of the work, and Section 5 defines the conclusion of the work.

2. Related Work

According to development, several techniques have been introduced to detect criminals in various scenarios and its states. Among them, a few recently developed work which is related to criminal detection is reviewed below.

Ratnaparkhi et al. [17] had designed an embedding method that maps facial features to a compact Euclidean face map, which can be used to detect differences in the face and a deep convolution neural network approach to detect criminals. This technique works well; however, it doesn't work well on multiple faces due to blurry or cropped images.

Sandhya et al. [18] had developed a system that uses a feed-forward neural network termed an autoencoder, whose input and output values are identical, to implement the deep learning neural network approach. One of the most well-known functions of auto-encoders is reconstructing the input image. As a result, use the encoder's output, which can be utilized for obtaining the input image by the decoder for identification verification rather than comparing the perpetrator's overall face. This technique uses a simple algorithm to assess how closely the input image equals images stored in the criminals' database. Although this method is efficient, it is time-consuming and overfitting inappropriate for multiple images.

Kumar et al. [19] had suggested a face recognition and criminal identification system utilizing a multi-task cascading network. This system will be capable of automatically distinguishing criminal faces. The system's one-shot learning mechanism would also require a single image of the criminal to identify. In this method, the criminal's face is recognized, the data for the identified criminal is gathered from the database, and a notification is sent to the police staff with every relevant detail and the position where the criminal was being watched by the camera. However, this method required a high level of time-consuming and model execution needs to be enhanced.

Venkatesh et al. [20] had evolved a classifier that used Fisher face and Local Binary Pattern Histograms (LBPH). It has been demonstrated that this technique works well for identifying criminal faces. The Fisher face classifier is useful for lowering dimensionality, whereas the LBPH classifier is known for its robustness to changes in illumination and subtle fluctuations in facial expressions. Combining these two techniques makes it possible to identify faces even in difficult lighting situations. However, this approach is not appropriate for rotating image prediction and consumes more time.

Ganji et al. [21] had used the well-known Principal Component Analysis method to develop an automatic facial recognition system for a criminal database. Automatic face detection and recognition will be possible with this system. In the absence of a thumbprint being present at the scene, this will aid law enforcement in identifying or detecting the criminal in the case. This technique works well, but it needs to be enhanced.

Amjad et al. [22] invented a facial recognition method based on deep learning that can identify or predict whether a person is a criminal or not and identify the probability they would be. A ResNet50 model that uses CNN and an SVM classifier to extract features from a dataset was utilized for training. This method is effective, but it has to be improved. Aherwadi et al. [23] had developed a method for face detection employing Face Encodings. In addition to providing the Police tremendous ease in identifying criminals, this version of the criminal detection system also saves them time because processes are automated. Although this strategy is efficient, it occasionally fails to identify appropriate people.

The above literature has numerous techniques and algorithms that were introduced to predict criminal. The outcome of these methods provide well performance but have some limitation that reduces the system's accuracy. The impacts that are having in the convolution model are reduced efficiency [21], pixel mismatch [17], overfitting [18], process consume more time as well as poor image quality [20], [22] and lack of robustness [19],[23]. To overwhelm these drawbacks, a novel method is proposed that enhances the facial image of criminal and non-criminal and improve accuracy. The next section presents a thorough explanation of the suggested methodology.

3. Proposed Methodology

Criminal identification and detection are slow and challenging processes. Nowadays, criminals are more clever than ever before, leaving no traces of biological evidence or fingerprints at the scene of the crime. Using advanced facial identification technology is a quick and simple solution. The majority of buildings and traffic lights now have CCTV cameras installed for surveillance purposes due to advancements in security technology. The camera's video footage can be utilized to identify suspects, criminals, runaways, and missing persons, among other things. These images are usually insufficient, making it challenging to recognize criminals and poor quality. Suitable enhancement methods are required for pre-processing the images. A multistage progressive V net image restoration technique is used in the proposed model to improve image quality. The proposed Multistage Progressive V-Net model process flow is depicted in Figure 1.

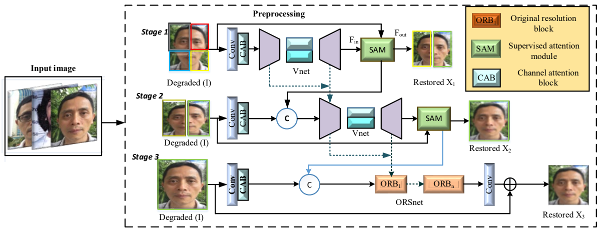

Figure 1. Architecture of proposed model.

The dataset used for this model includes both criminal and non-criminal images. Preprocessing the dataset is necessary due to the image's poor quality and difficulty in recognizing criminals. A Multistage progressive V-net based preprocessing technique is used in the proposed model. A three-stage CNN (convolutional neural network) called the multi-stage progressive V network (MPRV-Net) is used to restore images. For various types of Image restoration challenges such as deblurring, deraining and denoising, eliminated using an MPRV-Net and providing good performance on the dataset. The below section clearly explains the proposed methodology.

3.1. Dataset collection

The dataset used in this model was gathered from CK+ and fbi.gov websites which consist of both criminal and non-criminal images. To enhance the quality of the images as well as prediction, these datasets need to be preprocessed. The following section clearly explains the proposed preprocessing technique.

3.2. Preprocessing

Image restoration is a process of removing distortions like blur, noise, and rainfall from images to develop new pixels with clear details. For instance, the blur effect may be caused by the camera moving, the subject moving, the distortion of scattered light, the lack of proper depth of field, and the lens's softness. Noise is a term used to describe unwanted fingerprints and variations in the brightness or colour of the data. The unwanted effects of rainfall could impede the performance of the processing method. Therefore, it is crucial to employ pre-processing evaluators to eliminate these effects and improve the accuracy of the proposed model. The MPRV-Net-based preprocessing technique is used.

Multi-Stage Progressive V Net

Three steps combine to form MPRV-Net's progressive image restoration process. The first two stages are developed using a traditional V-net-based encoder-decoder model, which learns every contextual detail from the input image. The result uses the original resolution sub-network (ORSNet), which proceeds at the original image resolution resulting in spatially accurate results for pixel-by-pixel relation between the input and output images.

Include a supervised attention module between each stage rather than simply cascading several steps. The module rescales the feature maps from earlier stages before transferring them to the following stage, all under the surveillance of ground truth images. Introduce a method for integrating intermediate features from different stages of a network by using the intermediate multi-scale contextualized features of the earlier sub-network. The input image is accessed at each stage of the MPRV-Net stacking process. According to Fig. 1, the final result is divided into three distinct, non-overlapping stages: the initial phase using four patches, the second phase using two patches, and the final phase utilizing the original image. This division is typical of current restoration techniques. Similar to current restoration techniques, the multi-patch structure on the input image splits the image.

At each stage C, a diminished input image that I is included in the residual image R_C predicted by the proposed model, which produces X_C:

$$X_C = I + R_C \tag{1}$$

The following loss equation is used in the MPRV-Net's from start to finish optimization:

$$L = \sum_{C=1}^{3}\left[L_{char}(X_C, Z) + \lambda L_{edge}(X_C, Z)\right] \tag{2}$$

Where L_{char} is the char bonnier loss, and Z stands for the ground truth image.

$$L_{char} = \sqrt{\|X_C - Z\|^2 + \varepsilon^2} \tag{3}$$

With constant ε empirically adjusted to 10^{-3} for all experiments. Additionally, L_{edge} is the edge loss and is defined as:

$$L_{edge} = \sqrt{\|\Delta(X_C) - \Delta(Z)\|^2 + \varepsilon^2} \tag{4}$$

Where Δ represents the Laplacian operator, the relative relevance of the two loss factors is controlled by the value in equation (2), which is set to 0.05. The following section describes the key elements of the suggested method individually.

Key elements of MPRV-Net

The key element of the MPRV-Net model is,

- Processing of Complementary Features
- Cross-stage Feature Fusion
- Supervised Attention Module

In below section clearly explain each key feature,

- *Processing of Complementary Features*

Decoder-encoder and pipeline with only one scale of features are typically the architecture designs employed by single-stage CNNs for image restoration [24]. The input is gradually converted into low-resolution representations by the encoder-decoder networks, and after that, progressively reverse map and restore the original resolution. Although these models effectively encode multistate data, their frequent usage of down sampling techniques makes them susceptible to losing spatial characteristics. However, approaches that use a single-scale feature processing can be used to obtain images with excellent spatial features. However, the restricted receptive field makes their outputs less semantically robust. This demonstrates the inherent limitations of the aforementioned architectural design decisions, which can only have spatially accurate and contextually reliable effects. The following is a proposed multi-stage architecture that employs encoder-decoder networks in earlier stages and a network that works with the original input resolution in the final stage in order to utilize the advantages of both designs:

Encoder-Decoder Subnetwork: The encoder-decoder sub-network employs channel attention blocks (CABs) to acquire features at every stage. Figure 2 illustrates an encoder-decoder subnetwork developed using a conventional V net.

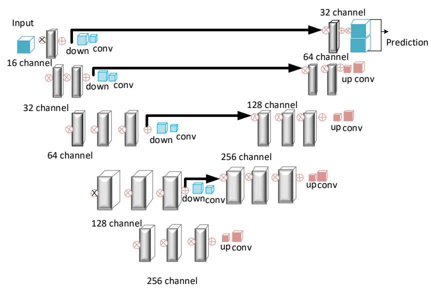

Figure 2. Architecture of V-net model.

The decompression path is positioned in the lower section of the V-Net, and the compression path is obtained in the higher section. A residual function is learned at each step of the calculation to speed it up, and it is then processed by ReLU before being added to the output of the final convolutional layer.

The bottom portion of the V-Net collects features and enhances the spatial support of lower resolution feature maps in order to gather and combine the necessary data and present volumetric segmentation. The foreground and background regions of the low- and high-resolution feature maps are probabilistically segmented using the soft-max function. Using a de-convolution process, the input complexity is increased. Features gathered from the initial stages of the compression path are transferred down to the bottom segment to improve the precision of contour prediction.

Using a combination of contour recovery and an approximation method, coordinates and the bounding box size are recovered and refined throughout the fusion stage. A distance ratio of 1:1 is used to determine whether two candidates are actually one discovery or two independent ones for candidates whose centres are too near to one another. The distance between the centres of two identified candidates is divided by the expected side of the bounding box from larger candidates to obtain the distance ratio. The results from each V-Net combine to determine whether the final forecast will be accurate. High-resolution features are computed by the MPR V-Net stage using a number of original resolution blocks (ORBs), each of which also contains CABs.

Original Resolution Subnetwork: Introduce the original-resolution subnetwork (ORSNet) in the final stage to maintain fine features from the input to the output image. While developing spatially enriched, high-resolution features, ORSNet does not use downsampling operations. There are several original-resolution blocks (ORBs) in it, each of those ORBs also contains CABs. ORSnet model effectively develops multiscale information, it has the ability to lose some spatial detail information by continuously down sampling. In the final stage, present the original resolution network (ORSNet) and include the enhancedd features from the input to the output image. Original resolution sub-network to generate rich spatial high-resolution features to make up for the loss of spatial information. At the network's end, the high-resolution rain line features obtained from the original resolution sub-network are combined with the original rain map to obtain the final rain removal image. ORSNet generates rich, high-resolution spatial features without down sampling operations, which consist of many original resolution blocks (ORBs). The structure of the ORB is shown in Figure 1,

- *Cross-stage Feature Fusion*

The CSFF module is integrated between two encoder-decoders and ORSNet in the proposed architecture. Instead of being transmitted to the following stage for aggregation, the characteristics from one stage are first improved by 11convolutions. The proposed CSFF has several advantages. Starting with the encoder-decoder, it continuously performs upward and down-sampling operations in order to reduce a network's sensitivity to information loss. Second, the multistate characteristics of one stage assist in enhancing the characteristics of the following stage. Third, by facilitating information flow, the network optimization process becomes more stable, allowing to expand the overall design by a number of steps.

- *Supervised Attention Module*

A clear image will be identified at each level by recent multi-stage image restoration networks, which is then passed on to the stage after that. To significantly boost performance, add a supervised attention module (SAM) between each pair of steps. The SAM schematic diagram is depicted in Figure 1 and details both of its contributions. For starters, it generates real-time supervisory signals useful at each progressive image restoration stage. Second, develop attention maps using locally supervised predictions that only allow the most efficient attributes to proceed to the following stage, suppressing the less valuable ones at this stage.

Figure 1 depicts the circumstances, the input parameters $F_{in} \in R^{H \times W \times Y}$ from the previous stage are processed by SAM, which then provides an unprocessed image $R_C \in R^{H \times W \times 3}$ using an automated 1×1convolution, where Y is the total number of channels, and $H \times W$ is the space dimension. The degraded input image I blended with the residual image resulting in the modified image $X_C \in R^{H \times W \times 3}$. Provide explicit supervision with the ground-truth image for the predicted image X_C.Using a 1×1 convolution and sigmoid activation, the image X_C is then used to create per-pixel attention masks$M \in R^{H \times W \times Y}$. Following the application of these masks to the modified local features, F_{in} provides attention-guided features that are subsequently included in the identity mapping path. Last but not least, problems with image restoration are eliminated by the attention enhanced representation F_{out} developed.

4. Result and Discussion

A novel Multistage Progressive V net for pixel restoration that significantly enhanced criminal images has been designed in this model. The objective manner of the dataset is collected from public websites divided into criminal and non-criminal categories for criminal recognition. These images are usually poor in quality and inadequate, making them difficult to recognize. As a result, it is crucial to employ pre-processing techniques to remove and restore the noise in the criminal image. A novel, multi-stage methodology that can result in outputs that are both contextually rich and spatially accurate is used in the proposed model. Due to its multi-stage structure, a framework is used to break down the difficult task of image restoration into easier steps and progressively restoration a diminished image. The proposed model performance is analyzed using Intel Core i7 CPU, NVIDIA GeForce RTX 3070 GPU, and 64GB RAM software.

4.1. Dataset Description

In the designed model, the dataset is collected through CK+ and fbi.gov websites [25] [26]. Overall, 934 image data are collected, split into criminal and non-criminal groups. In this group, 339 images are criminal, and 595 images are non-criminal considered.

The images are frequently insufficient, poor quality and difficult to identify. For pre-processing the criminal and non-criminal images, suitable enhancing techniques are required [27]. The proposed model enhances the image quality using a multistage progressive v net image restoration technique. This algorithm consists of 3 stages, which split the image and learn every feature using the V-net architectures. Figure 3 represents pre-processing results of the proposed dataset [28].

Original image	**Preprocessing image**
	

Figure 3. The outcome of the proposed model pre-processing.

MPRV-net image performance metrics are compared with a few other current models in the following part to analyze image quality. Average gradient, mean square error (MSE), root mean square error, peak signal-to-noise ratio, signal-to-noise ratio, and correlation coefficient are the metrics used for comparison. Each metric is individually analyzed, and the values are observed to analyze the performance of the pre-processing approach.

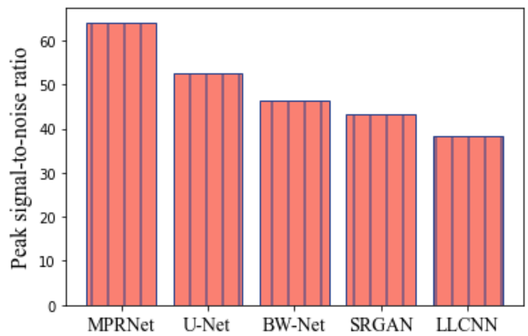

Figure 4. Peak signal-to-noise ratio for the proposed and existing model.

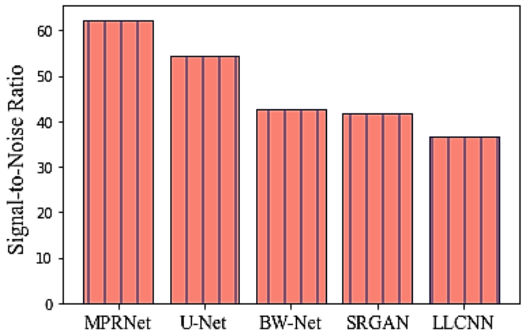

Figure 5. Signal-to-noise ratio for the proposed and existing model.

A comparison of the peak signal-to-noise ratio (PSNR) and signal-to-noise ratio (SNR) of the proposed MPRV-net model provides for testing the methodology shown in figure 4 and figure 5. The PSNR is the ratio of a signal's highest practicable value (power) to the highest possible amount of distorted noise that compromises the accuracy of its representation. The proposed MPRV-net model provides 64.1 PSNR, U net have 52.6 PSNR, BW-Net have 46.3 PSNR, SRGAN have 43.2 PSNR and LLCNN model have 38.4 PSNR. Thus, the proposed model provides a higher PSNR value than the traditional approaches shown in figure4. Then the signal-to-noise ratio is analyzed, and its comparison is shown in figure5. The SNR will rise with a larger field of vision because a larger pixel size results in more signals being received by each individual pixel. Large pixels will pick up more signals, resulting in high SNR images. The proposed MPRV-net model provides 62.3 SNR, Unet have 54.3 SNR, BW-Net have 42.8 SNR, SRGAN have 41.7and LLCNN model have 36.6 SNR. It demonstrates the proposed model provides a better SNR value than another models.

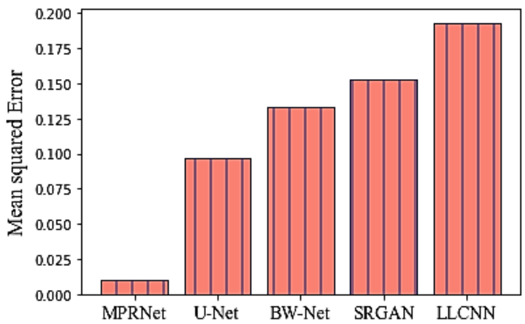

Figure 6. Mean square error for the proposed and existing model.

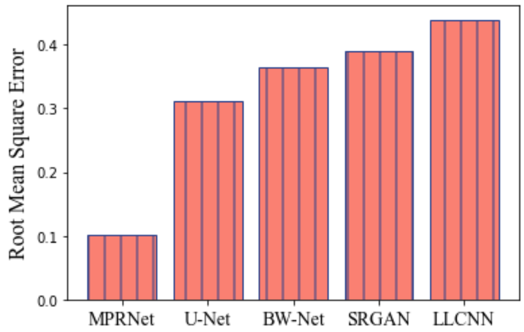

Figure 7. Root mean square error for the proposed and existing model.

The proposed MPRV-net, MSE and RMSE are compared with the methods that are already in use, as shown in figure 6 and figure 7. To validate the procedure that was illustrated in figure 6, the MSE is analyzed, and the observed values are compared to the convolution model. The proposed MPRV-net model has 0.0105MSE, U net has 0.0968MSE, BW-Net has 0.1332MSE, SRGAN has 0.1524MSE, and the LLCNN model has 0.1932MSE. Thus, show that the proposed model has a lower error value than the traditional approaches. The method depicted in figure 7 is then validated by doing an MSE analysis and comparing the observed values to the convolution model. The proposed MPRV-net model has 0.10246RMSE, whereas U net, BW-Net, SRGAN, 0.3649RMSE, and the LLCNN model each have 0.39038RMSE, 0.3649RMSE, and 0.31112RMSE, respectively.

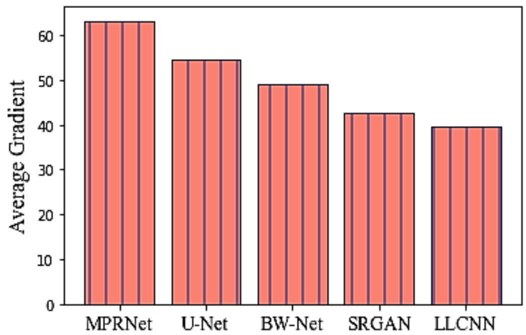

Figure 8. Average gradient for the proposed and existing model.

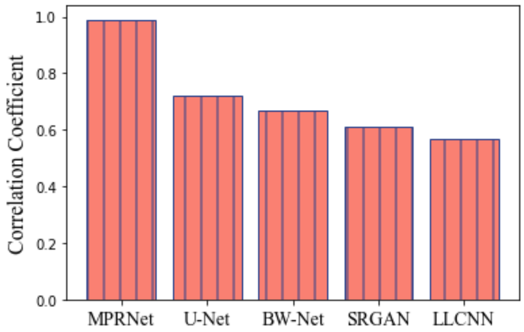

Figure 9. Correlation coefficient for the proposed and existing model.

The proposed MPRV-Net average gradient and correlation coefficient are compared with the existing methods, as shown in Figures 8 and 9. The average gradient develops the relationship between the image's clarity and the pattern's finely detailed variance. The proposed MPRV-net model provides a gradient of 63.2, whereas Unet provides a gradient of 54.6, BW-Net a gradient of 49.2, SRGAN a gradient of 42.8, and the LLCNN model provides a gradient of 39.5. It indicates that when compared to existing approaches, the proposed MPRV-net model has high gradient values in figure 8. Figure 9 shows the analysis of the correlation of the image. When performing correlation, a filter mask, frequently referred to as a kernel, is moved over the image, and the sum of the products is calculated at each place. The proposed MPRV-net model provides 0.99 correlation, Unet have 0.72 correlation, BW-Net have 0.67 correlation, SRGAN have 0.61 correlation, and the LLCNN model have 0.57 correlation. It demonstrates that the proposed model provides a better average gradient and correlation coefficient value than other models.

Confusion matrix

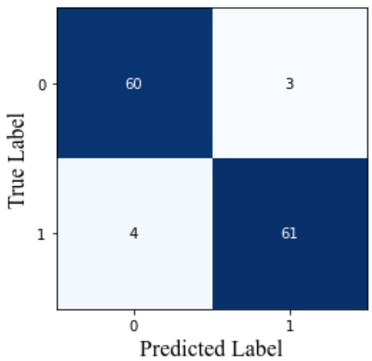

Figure 10. Confusion matrix of the proposed method.

Figure 10 shows the proposed model confusion matrix model. A classification algorithm performance is determined using a confusion matrix technique, which is a collection of data that is organized according to both actual and expected data. For the performance analysis, data obtained for such a system are evaluated. The predictive analysis technique provides a confusion matrix that includes positive and negative rates (true and false). Figure 10 makes it clear that 61 and 60 images are correctly predicted in 0 and 1 class, respectively, whereas 3 and 4 images are wrongly predicted. The positive and negative rates are also used to measure the accuracy, sensitivity, specificity, and null error rates. The performance metrics are compared to existing approaches like U Net, BW-Net, SRGAN and LLCNN models.

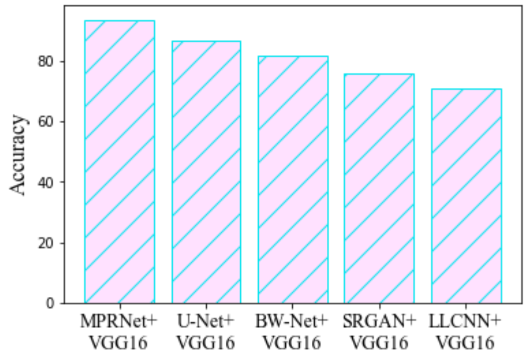

Figure 11. Comparison of accuracy for the proposed and existing model.

Figure 11 shows the accuracy values, which were calculated using the confusion matrix. ADD some points about performance metrics and accuracy. The system that predicts a value with the least degree of error is said to be accurate. The proposed method has a 94% accuracy rate, compared to Unet at 87%, BW-Net at 82%, SRGAN at 76%, and LLCNN at 71%.

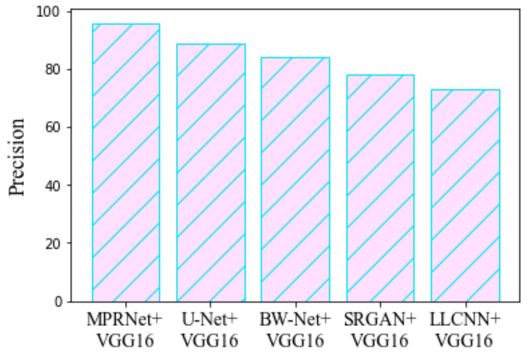

Figure 12. Comparison of precision for the proposed and existing model.

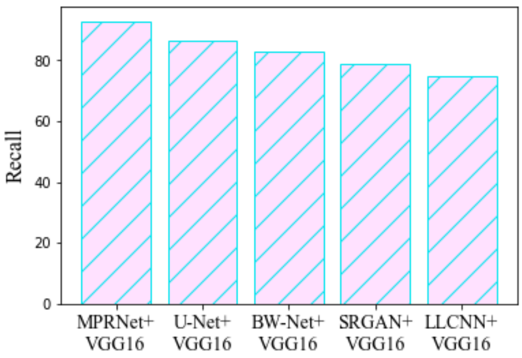

Figure 13. Comparison of recall for proposed and existing model.

A comparison of the intended and existing approaches' precision is shown in Figure 12. The quantity of anticipated favourable events is what is meant by accurate measurement. In comparison to other existing approaches, such as Unet, BW-Net, SRGAN and LLCNN, with corresponding precision values of 89%, 84%, 78%, and 73%, the proposed method's precision value was discovered to be 96%. As shown in figure 13, recall is calculated by dividing the total number of components that actually fall into the positive class by the number of true positives. In comparison to other existing approaches, such as Unet, BW-Net, SRGAN and LLCNN, with corresponding recall values of 86.4%, 83%, 79% and 75%, the proposed method's recall value was discovered to be 93%.

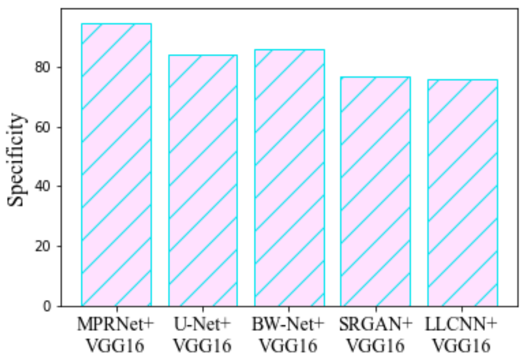

Figure 14. Comparison of specificity for proposed and existing model.

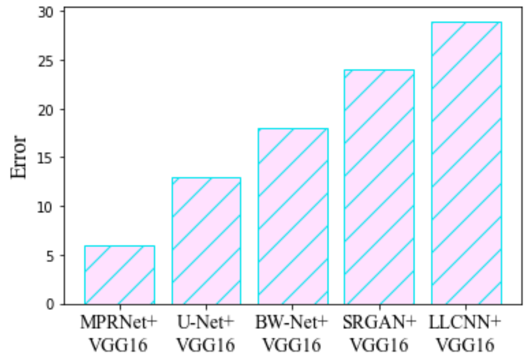

Figure 15. Comparison of error for the proposed and existing model.

Figure 14 shows a comparison of the specificity of the proposed and current techniques. The degree to which a model can predict the real negatives of every imaginable sort is known as specificity. In comparison to various approaches, such as Unet, BW-Net, SRGAN and LLCNN, with corresponding specificity values of 84%, 86%, 77% and 76%, the proposed method's specificity value was determined to be 95%. The system operates worse when the error is high and better when the error is low. Figure 15 illustrates the error contrast among the proposed and existing methodologies. The proposed method's error rate is 6%, Unet's is 13%, BW-Net's is 18%, SRGAN's is 24% and LLCNN's is 25%. It demonstrates that the proposed model has lower error rates than the conventional models.

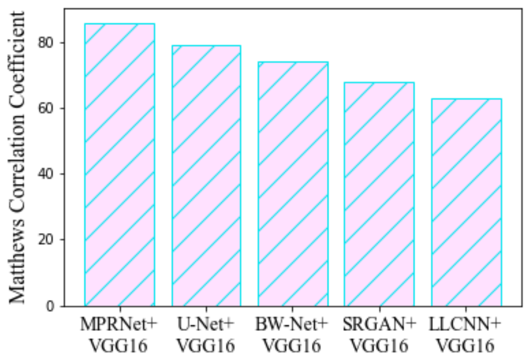

Figure 16. Comparison of Matthews's correlation coefficient for the proposed and existing model.

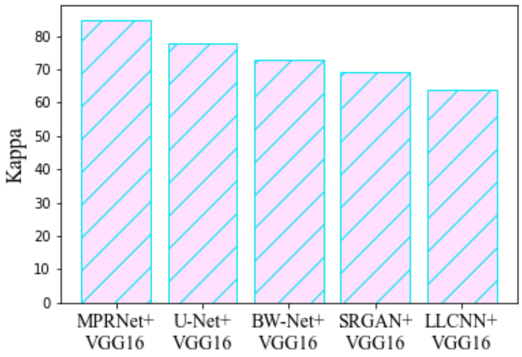

Figure 17. Comparison of kappa for the proposed and existing model.

Figure 16 illustrates an MCC comparison of the proposed and existing methods. The purpose of MCC is to evaluate or quantify the discrepancy between expected and actual values. MCC in the proposed method has a value of 86%, Unet of 79%, BW-Net of 74%, SRGAN of 68% and LLCNN of 63%. Kappa comparison of the suggested and existing approaches is shown in Figure 17. A statistical indicator of the consistency of different variables across rates is called kappa. The proposed method's kappa value is 85%, Unet is 78%, BW-Net is 73%, SRGAN is 68%and LLCNN is 64%.

4.2. Time comparison analysis

In this section, the overall working process time comparison was provided. Process Cycle Time refers to the duration of the overall process. It may be computed instantly by adding the total sum of time spent on each process step and the delay time. Figure 18 and figure 19, 20 shows the training time, testing time, and execution time. The numbers are compared to Unet, BW-Net , SRGAN and LLCNN models to confirm the procedure.

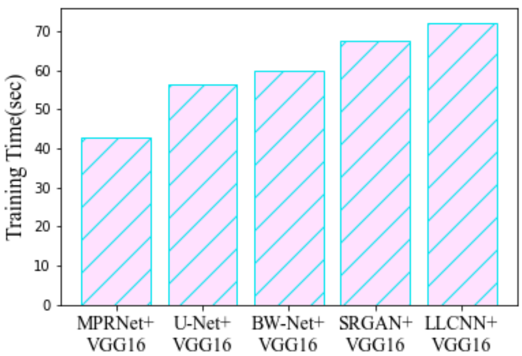

Figure 18. Analysis of training time for the proposed and existing model.

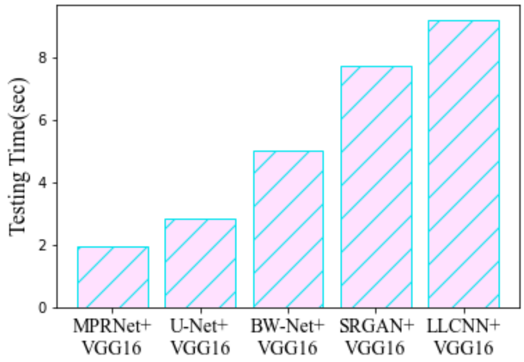

Figure 19. Analysis of testing time for the proposed and existing model.

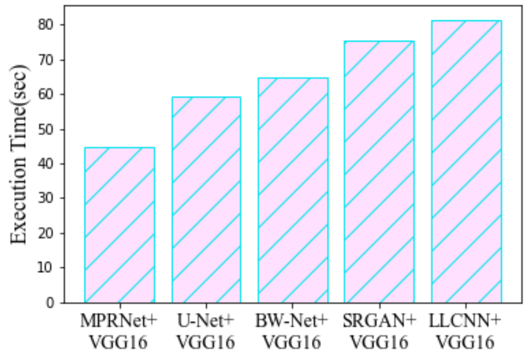

Figure 20. Analysis of execution time for the proposed and existing model.

During the training period, 80% of the data were used to train the MPRV-net model for effective prediction, as the existing models were also trained using 80% data. The training period causes some time to complete the training process. In figure 18, the training time of the proposed model is 42.72 ms, the Unet model consumes 35 ms, the BW-Net model takes 41 ms, the SRGAN model consumes 45 ms, and the LLCNN consume 50 ms. The testing time is also analyzed and plotted in figure 19. It demonstrates that the proposed MPRV-Net takes 1.96 ms, the existing Unet model consumes 2.84 ms, the BW-Net model takes 5.02 ms, SRGAN model takes 7.73 ms, and the LLCNN model takes 79.21 ms. Then the proposed model overall execution time is analyzed and compared to some other models that are shown in figure 20. The overall execution time is termed the addition of training and testing time. The proposed model consumes 44.68 ms, the Unet model consumes 59.14 ms, the BW-Net model takes 64.82 ms, the SRGAN model consumes 75.34, and the LLCNN consume 81.51 ms to complete the process. The above comparison demonstrates that the proposed model effectively improves the prediction accuracy of any input image.

5. Conclusion

A multistage progressive V-net image restoration technique is used in the proposed model to improve image quality. Performance is limited by unwanted traces and fluctuations in the images' brightness or colour, making recognition challenging. So, using a revolutionary image restoration technique, it is proposed to eliminate distortions from images like blur, noise, and rainfall in order to develop new, clear ones. At first, the criminal images with different styles of each criminals' and non-criminals' datasets are collected through public websites. The multi-stage progressive image restoration network (MPRV-Net) has three steps for progressively restoring images. The input criminal and non-criminal images have been employed in the first two steps to develop an encoder-decoder model based on a common V-net. The original resolution sub-network (ORSNet) generates spatially appropriate results for a pixel-to-pixel consistency between the input and output images in the final stage at the original image resolution. Because of their hierarchical multi-scale representation and computational effectiveness, encoder-decoder-based V-net architectures have been mostly used for recovery in convolutional networks. MPRV-Net revealed an enhanced multi-layer architecture developed to accommodate high-level global characteristics and local specifics. Channel attention blocks (CABs) extract the encoder-decoder sub-networks characteristics at each level, followed by bi-linear interpolation and a convolution layer. Multiple original resolution blocks (ORBs), each including CABs, are used to compute detailed features. A cross-stage feature fusion (CSFF) module is implemented between each pair of stages. As a result of multiple up- and down-sampling, the network as a whole becomes more robust and stable, besides improving information flow.

Additionally, the efficacy of the proposed MPRV-net approach was analyzed and evaluated against other conventional approaches like U-net, BW-net, SRGAN and LLCNN. The proposed approach has 94% accuracy, 93% recall, 95% specificity, 86% MCC and 85% kappa. The comparative analysis demonstrates that the suggested model provides a more efficient result than the present approaches. In future work, the enhanced images are used for detecting the facial landmarks even with different sizes, lighting and poor pixel quality that effectively predict the criminal face using neural network.

Acknowledgements.

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of Interest: The authors declared that they have no conflicts of interest in this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Availability of data and material: Not applicable

Code availability: Not applicable

Author contributions: The corresponding author claims the major contribution of the paper including formulation, analysis and editing. The co-author provides guidance to verify the analysis result and manuscript editing.

Compliance with ethical standards: This article is a completely original work of its authors; it has not been published before and will not be sent to other publications until the journal's editorial board decides not to accept it for publication.

Reference

[1] Apoorva P, Impana HC, Siri SL, Varshitha MR, Ramesh B. Automated criminal identification by face recognition using open computer vision classifiers. In2019 3rd International Conference on Computing Methodologies and Communication (ICCMC) 2019 Mar 27 (pp. 775-778). IEEE.

[2] Hasan BM, Abdulazeez AM. A review of principal component analysis algorithm for dimensionality reduction. Journal of Soft Computing and Data Mining. 2021 Apr 15;2(1):20-30.

[3] Payal P, Goyani MM. A comprehensive study on face recognition: methods and challenges. The Imaging Science Journal. 2020 Feb 17;68(2):114-27.

[4] Porcu S, Floris A, Atzori L. Evaluation of data augmentation techniques for facial expression recognition systems. Electronics. 2020 Nov 11;9(11):1892.

[5] Kortli Y, Jridi M, Al Falou A, Atri M. Face recognition systems: A survey. Sensors. 2020 Jan 7;20(2):342.

[6] Sun S. Application of fuzzy image restoration in criminal investigation. Journal of Visual Communication and Image Representation. 2020 Aug 1;71:102704.

[7] Liu H, Cao F, Wen C, Zhang Q. Lightweight multi-scale residual networks with attention for image super-resolution. Knowledge-Based Systems. 2020 Sep 5;203:106103.

[8] Lei X, Pan H, Huang X. A dilated CNN model for image classification. IEEE Access. 2019 Jul 8;7:124087-95.

[9] Amirian S, Rasheed K, Taha TR, Arabnia HR. Automatic image and video caption generation with deep learning: A concise review and algorithmic overlap. IEEE access. 2020 Dec 4;8:218386-400.

[10] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image restoration. IEEE transactions on pattern analysis and machine intelligence. 2020 Jan 22;43(7):2480-95.

[11] Brooks T, Mildenhall B, Xue T, Chen J, Sharlet D, Barron JT. Unprocessing images for learned raw denoising. InProceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2019 (pp. 11036-11045).

[12] Zhang H, Sindagi V, Patel VM. Image de-raining using a conditional generative adversarial network. IEEE transactions on circuits and systems for video technology. 2019 Jun 3;30(11):3943-56.

[13] Suin M, Purohit K, Rajagopalan AN. Spatially-attentive patch-hierarchical network for adaptive motion deblurring. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2020 (pp. 3606-3615).

[14] Ren D, Zuo W, Hu Q, Zhu P, Meng D. Progressive image deraining networks: A better and simpler baseline. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 3937-3946).

[15] Tao X, Gao H, Shen X, Wang J, Jia J. Scale-recurrent network for deep image deblurring. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 8174-8182).

[16] Zamir SW, Arora A, Khan S, Hayat M, Khan FS, Yang MH, Shao L. Multi-stage progressive image restoration. InProceedings of the IEEE/CVF conference on computer vision and pattern recognition 2021 (pp. 14821-14831).
[17] Ratnaparkhi ST, Tandasi A, Saraswat S. Face detection and recognition for criminal identification system. In2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) 2021 Jan 28 (pp. 773-777). IEEE.
[18] Sandhya S, Balasundaram A, Shaik A. Deep Learning Based Face Detection and Identification of Criminal Suspects. Computers, Materials & Continua. 2023 Feb 1;74(2).
[19] Kumar KK, Kasiviswanadham Y, Indira DV, Bhargavi CV. Criminal face identification system using deep learning algorithm multi-task cascade neural network (MTCNN). Materials Today: Proceedings. 2023 Jan 1;80:2406-10.
[20] Venkatesh M, Dhanalakshmi C, Adapa A, Manzoor M, Anvesh K. Criminal Face Detection System.
[21] Ganji A, Kamat R, Biradar SC, Gadivaddar S. Criminal Face Detection Using Machine Learning.
[22] Amjad K, Malik AA, Mehta S. A technique and architectural design for criminal detection based on lombroso theory using deep learning. Lahore Garrison University Research Journal of Computer Science and Information Technology. 2020 Sep 25;4(3):47-63.
[23] Aherwadi NB, Chokshi D, Pande DS, Khamparia A. Criminal Identification System using Facial Recognition. InProceedings of the International Conference on Innovative Computing & Communication (ICICC) 2021 Jul 12.
[24] Haque IR, Neubert J. Deep learning approaches to biomedical image segmentation. Informatics in Medicine Unlocked. 2020 Jan 1;18:100297.
[25] Dataset 1: https://www.fbi.gov/wanted/fugitives
[26] Dataset 2: https://paperswithcode.com/dataset/ck
[27] Ratnaparkhi ST, Singh P, Tandasi A, Sindhwani N. Comparative analysis of classifiers for criminal identification system using face recognition. In2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) 2021 Sep 3 (pp. 1-6). IEEE.
[28] Jhanani R, Harshitha S, Kalaichelvi T, Subedha V. Mobile Application for Human Facial Recognition to Identify Criminals and Missing People Using TensorFlow. Journal of Research in Engineering, Science and Management. 2020;3(4):16-20.