International Journal of Applied Engineering & Technology

PREPROCESSING OF IMAGES CAPTURED FROM CCTV VIDEO USING FILTERING TECHNIQUES

[1]S. Nishanthi and [2]Dr. S. Santhosh Baboo
[1]Research Scholar and [2]Principal & Supervisor, PG and Research Department of Computer Science and Applications, Dwaraka Doss Goverdhan Doss Vaishnav College, Arumbakkam, Chennai, India.
[1]nishanthis610@gmail.com and [2]santhos1968@gmail.com

ABSTRACT
Image processing is a branch of signal processing that typically uses computer-aided processing. Medical imaging, remote sensing, computer vision, video processing, and multimedia systems are all examples of applications where image processing is employed extensively. In the context of digital imaging, video processing refers to the application of various techniques to a succession of images (frames) to evaluate, improve, or modify video data. The CCTV For the preprocessing technique, video is taken and converted into frames, and then the cascade object detector uses the Viola-Jones method to identify faces, noses, eyes, mouths, and upper bodies. After face detection, the face is extracted and converted into grayscale, resizing image, noisy image, Median Filter, and Weiner Filter techniques are used. The Existing Algorithm uses the SRCNN and VDSR algorithms from a deep learning-based approach for image super-resolution. The image's performance is calculated using the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) to assess the quality of reconstructed or enhanced images.

Keywords: Super-Resolution Convolutional Neural Network, Very Deep Convolutional Networks, Peak Signal-to-Noise Ratio and Structural Similarity Index.

1. INTRODUCTION
Digital Image Processing deals with applying various algorithms to manipulate digital images. For many applications, including image compression, object detection, and face recognition, it is a necessary preprocessing step. Image processing can be used to enhance an image's quality, eliminate unwanted components from an image, or even begin from scratch and create new images. Image processing, for instance, can be used to eliminate the background from a portrait picture so that the subject is only visible in the forefront. Image enhancement, or increasing the quality of an image, is one of the most typical image processing operations. It is useful in computer vision tasks, remote sensing, and surveillance. Adjusting the image's contrast and brightness is a typical strategy. The contrast of an image is the difference in brightness between its brightest and darkest portions. The overall brightness of an image can be improved by increasing the contrast, making it easier to perceive. The brightness of an image refers to its overall lightness or blackness. An image can be made lighter by increasing the brightness, making it easier to view. Most image editing tools can automatically alter contrast and brightness, or they can be adjusted manually.

Image Restoration is particularly important since advanced methods in this field have the ability to recover damaged historical materials. Deep Learning-based image restoration algorithms may be able to recover huge sections of lost information from damaged documents. The technique of converting a single picture into many segments or regions is known as image segmentation. Image segmentation is frequently employed as a preprocessing step for object detection since each segment represents a different object in the image. There are numerous picture segmentation algorithms available, but one of the more prevalent techniques is to use thresholding. For example, binary thresholding is the process of turning an image into a binary image in which each pixel is either black or white. The threshold value is chosen so that all pixels with brightness levels lower than the threshold are black, and all pixels with brightness levels greater than the threshold are white.

Jaiswal *et al*. [1], this research paper tells that an artificial neural network-based solution that will function on CCTV video sequences. Secondly, proposed digital image processing techniques for moving object feature extraction, and finally, used concepts of artificial neural network and digital image processing techniques to detect moving objects in bad weather conditions such as fog, rain, evening time, night time, and light luminosity. The projected experimental findings will indicate that when utilizing the suggested approach to detect moving objects in video sequences, the accuracy can reach 99% and the calculation speed is very quick, which might meet the demand for real-time video processing. An exploration study was finished by Ullah Rehmat *et al*. [2] in their work by the principal component analysis (PCA) and convolutional neural network (CNN) feature extraction algorithms, as well as evaluating the performance of the algorithms K-nearest neighbor (KNN), decision tree, random forest, and CNN. For simulation and performance evaluation, these algorithms are applied to a dataset of over 40K real-time photos recorded at various parameters such as light level, rotation, and scale. Finally, were able to recognize faces with less computing time and greater than 90% accuracy.

Kurniawan *et al*. [3] introduces an image classification approach, a convolutional neural network is used to predict traffic congestion. On the CCTV camera monitoring collected images dataset, the CNN model attained an average accuracy of 89.50%, just resize and convert the photos to 100x100 grayscale images. First, the standard technique of crowd analysis through several qualities is explained by Bansod *et al*. [4], the CNN-based crowd-analysis models are explored. Finally, demonstrated crowd analysis utilizing the CNN object model and various RNN training methods. The use of Inception Convolution Networks to extract features, followed by a single-layer LSTM RNN, resulted in high video classification accuracy. Vinay. A *et al.* [5], has shown Forensic analysis is one of the most useful tools in the investigation of crime. To aid in the investigation of criminal cases, one way is to improve the visual quality of CCTV video. It is used to overcome the issue of not having sufficient funds to purchase a high-quality image/video capture gear. Pure image processing approaches and machine learning techniques are used to address the challenge of image enhancement.

Accidents are a leading cause of death in India. Over eighty percent of accident-related deaths are caused not by the event itself, but by a lack of timely assistance to accident victims is discussed by Smitha K *et al*. [6], and the concept is to run each frame of a video through a deep learning convolution neural network model that has been trained to classify video frames as accident or non-accident. Convolutional Neural Networks have proven to be a quick and accurate method of picture classification. CNN-based image classifiers have achieved more than 95% accuracy. Elhoseny *et al*. [7] has developed, The Kalman filtering technique is used to monitor moving objects in video frames. After differentiating the items, the probability-based grasshopper algorithm was used to optimize parameters using Kalman filtering. A similarity measure was used to track the selected items in each frame using the best parameters. Finally, the proposed MODT framework was put into action, and the results were evaluated. According to the results of the studies, the MODT framework obtained maximum detection and tracking accuracies of 76.23% and 86.78%, respectively.

An exploration work by Bai *et al*. [8], SRCNN was developed using accelerated bicubic interpolation and tested with 2000 digital rock core pictures. The experiment demonstrated that the accelerated bicubic interpolation algorithm performed better than the region-based bicubic image interpolation algorithm and the standard bicubic interpolation algorithm, as well as the feasibility of using SRCNN based on the algorithm it suggested to generate higher resolution digital rock core images. Another work did by Ooi *et al*. [9] in their research work, The super-resolution convolutional neural network (SRCNN) was the first CNN-based algorithm, and it is still being improved today using various methodologies. The methodologies used included the type of loss functions used, the upsampling module applied, and the network design strategies used.

An experimental work did by Mohan Amrita *et al*. [10], tells from the SRCNN approach is based on a deep convolutional neural network and is a rapidly evolving topic with several practical applications in various fields. SRCNN's main goal is to create a high-resolution image from low-resolution landslide images. The proposed algorithm's performance on landslide images is meaningfully evaluated using the Peak Signal to Noise Ratio (PSNR), Mean Squared Error (MSE), and Structural Similarity (SSIM) index. Elsaid *et al*. [11], has studied the efficacy of applying an innovative deep-learning method, super-resolution convolutional neural network (SRCNN), to create sub millimeter super-resolution diffusion-weighted (DW) pictures was validated in this study.

The 2D-based deep-learning algorithm was validated using numerical simulations and by investigating region-of-interest (ROI) using real human data from three healthy participants.

Zhangzong Zhao *et al*. [12] has suggested method, one of the most significant super-resolution algorithms is SRCNN. First, immediately parallelize and build the SRCNN, then accelerate the convolution by utilizing the GPU memory's hierarchical feature. Experiments demonstrate that the entire execution time for 1080p to 4K upscaling is decreased from 300s per frame to 0.15s per frame while maintaining the same image quality as the original SRCNN. An experimental work did by Vadim Romanuke *et al*. [13] , the best VDSR network outperforms bicubic interpolation by over 3.2%. VDSR-14/0.95, VDSR-20/0.95, and VDSR-16p4 networks all exceed it by at least 3%. In terms of performance, both networks are nearly identical. Despite the fact that VDSR-14/0.95 is taught faster, the performance (as well as the benefit over bicubic interpolation) of VDSR-20/0.95 and VDSR-16p4 improves as training advances. VDSR-14 is the most quickly trained and still has a slight advantage.

Donghyeon Lee *et al*. [14], the very deep convolutional network for image super-resolution (VDSR) technique is a promising SISR approach, but it is too complicated to implement in hardware for commercial goods. The suggested approach intends to implement VDSR with relatively few hardware resources while limiting image quality degradation exhibited in simulation results. The 1D reorganization reduces the number of multiplies by 55.6%, while the 1D vertical filter size reduces half the buffer size. As a result, the suggested system processes a full-HD video in real-time with 8,143.5k gates and 333.1kB SRAM, however, the image quality suffers a 1.06dB degradation when compared to VDSR. Reddy, M *et al.* [15], has deals to improve the resolution of photos recorded using underwater applications, the Very-Deep Super-Resolution (VDSR) reconstruction model is introduced. This research introduces a residual learning model for underwater image improvement. In order to generate the feature map, the CNN layers are calculated and applied to the images.

Another exploration work completed by Mehmet Cem Catalbas *et al*. [16], in their experimental work. The VDSR is one of the most widely used CNN-based super-resolution algorithms, and this research attempts to improve the performance of the VDSR network structure's output by employing the natural image quality evaluator (NIQE). When combining the residual picture derived from the VDSR approach and the original low-resolution image, new coefficients have been proposed to increase performance. D. Vint *et al*. [17] in their research work, VDSR Single Image Super Resolution architecture is used to boost the spatial resolution of low-resolution photographs. Two sets of tests are carried out to achieve this goal. The former was performed on real-world photos to assess the network's capacity to improve low-resolution photographs. The second test is synthetic and is performed on images of a resolution chart. This demonstrated that the network can improve the spatial resolution of provided images, which is useful when images have been affected by conditions like diffraction. Donghyeon Lee *et al.* [18], proposed in terms of image quality, VDSR beats the CNN-based approaches. However, it necessitates a high level of computer complexity, which prohibits real-time processing. To reduce computational complexity, the method of applying a deconvolution layer to VDSR is proposed in this study. In comparison to the original VDSR, the proposed approach produces a 4.46 times speed-up with a minor reduction in image quality of less than -0.1 dB.

2. MATERIALS AND METHODS

This research aims to improve video resolution by employing an innovative technique for video processing. The study used a collection of low-resolution images gathered from various sources as the baseline data for experimentation. Preprocessing the videos into frame conversion to reduce noise and filtering technique was part of the process. Following that, an innovative upscaling method was used to interpolate pixel information and boost video resolution. This algorithm is used for Existing methods which are Super-Resolution Convolutional Neural Network (SRCNN) and Very Deep Super Resolution (VDSR), which were critical in transforming low-resolution frames into high-resolution frames. A variety of quantitative indicators, including Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) were used to assess the performance of the suggested technique**.** The dataset taken from Kaggle dataset.

2.1 Data Set

The Dataset used in this study was obtained from Kaggle, a popular site for open datasets. The dataset contains a CCTV videos footage for this research work. This work has in excess of 3000 videos and which took the videos of 400 for performance evaluation. Figure 1 shows the sample Dataset of an CCTV video.

Figure 1: Sample Dataset

2.2 Preprocessing

Preprocessing is plays a vital role in digital image processing for improving image quality and retrieving important information. Following that, frame conversion is used, and the best frame from that conversion is used to recognize people's faces, noses, eyes, lips, or upper body using the Viola-Jones algorithm. Grayscale pictures are frequently employed as a preprocessing step in a variety of image recognition tasks. By displaying images in shades of gray rather than full-color RGB values, grayscale decreases image complexity. Image quality was improved using noise reduction techniques such as Median filtering and Weiner filtering. The architecture of filtering technique is shown in Figure 2.

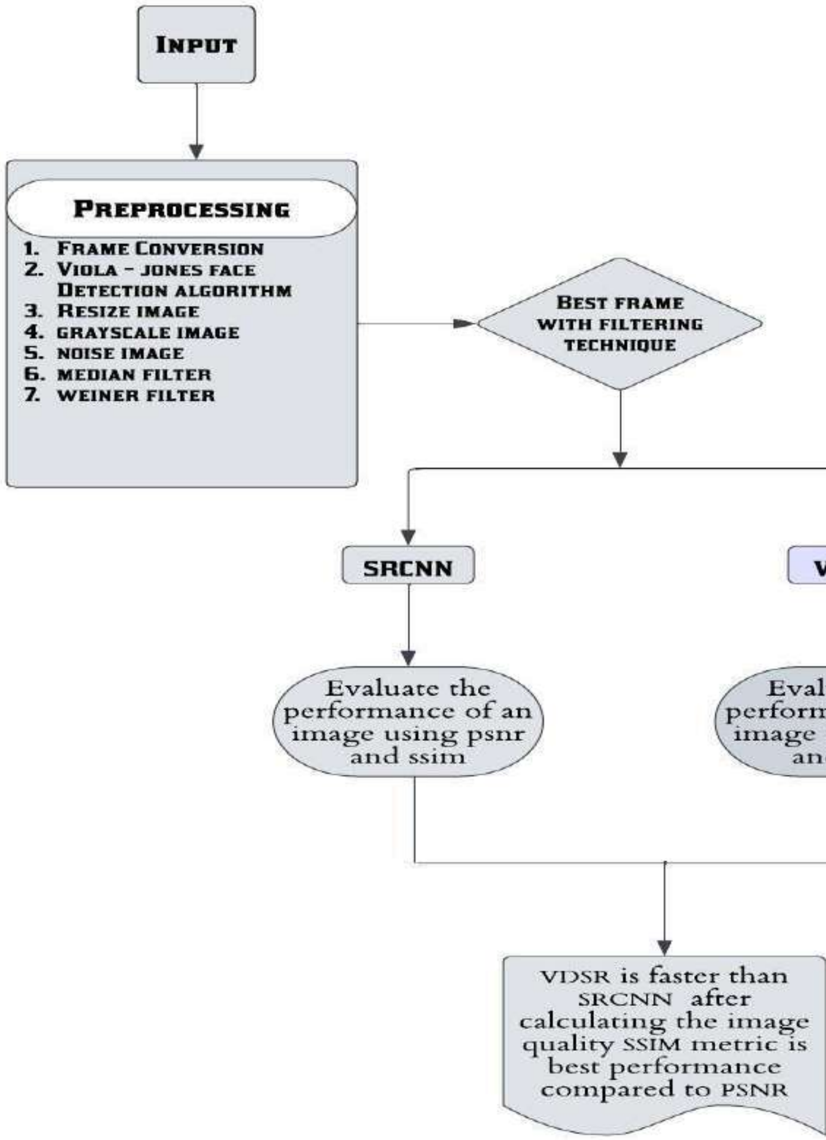

Figure 2: The architecture of filtering technique

1. Frame Conversion

Chandra Shekhar Mithlesh *et al*. [19], Frame conversion is used to extract frames from that CCTV videos. Every frame F is determined by the video structure. The video sequence is divided into frames, which are taken one by one from the video structure. The larger the size, the more frames are calculated. Video frame extraction aids in surveillance system security, change detection systems, video summarization, indexing, online video processing, and other applications. The video is in AVI (Audio Video Interleave) format. And the frame is saved in the jpg format.

2. Viola-Jones Face objects Detection Algorithm

Viola-Jones Face Detection Algorithm was the first to detect faces in real-time. The primary goal is to create fully automated human facial measuring systems using images with complex backdrops. The detection of facial characteristics such as the eye, nose, and mouth is an essential phase in many subsequent facial image processing tasks. The primary study of face detection is to detect the section of the body and mention the circle or rectangular of each portion of the body. Low-level and high-level analysis, feature analysis, and active shape model analysis are all subsets of face feature-based techniques is explained by Alpika Gupta *et al.* [20].

3. Resize Image

After face extract from the face detection resized is applied for the image. Resizing an image is a technique used in digital image processing that involves modifying the image's dimensions. Image compression, content delivery, and preparing images for varied display or processing requirements are all prominent uses for resizing.

4. Grayscale Image

After resizing Grayscale image is applied which is used for it removes all color information and simply leaves several shades of gray, with the brightest being white and the deepest being black. Each pixel is a representation of the image's luminous intensity. Grayscale is the most basic model because it describes colors with only one component, lightness. A value ranging from 0 (black) to 255 (white) is used to define the amount of brightness.

5. Noise Image

Salt-and-pepper noise is a sort of digital picture noise that appears as a random distribution of white and black pixels in an image. It has the look of salt and pepper grains. This form of noise can be created by a variety of sources, such as transmission mistakes, sensor malfunctions, or picture acquisition process flaws. To generate a salt-and-pepper noise image, change the intensity values of specific pixels at random to either the maximum or minimum value allowed in the image.

6. Median Filter

Rong Zhu et al. [21], the median filter is non-linear filter. The most used technique for removing image noise is the median filter. Without processing the pixels designated as signals, the algorithm detects image noise and creates a noise-marked matrix based on the properties of salt and pepper noise. It is demonstrated that an enhanced median filter denoising technique may significantly cut down on program execution time while maintaining a higher level of image detail, making it more appropriate for standard computer image denoising. There are multiple uses for median filters in image processing software. However, median filters tend to remove fine details such as lines and sharp edges, fail to efficiently eliminate heavy-tailed noise, and perform poorly when signal-dependent noise is included.

7. Weiner Filter

Iman Hussein AL-Qinani [22], the Weiner filter is a linear filter. It is ideal for removing noise and motion blur from images. This technique produces an image with less noise than the original image. The Wiener filter is the most effective method for removing blur in images caused by unfocused optics or linear motion. Weiner filters are by far the most used deblurring approach since they mathematically determine the optimal output. The Weiner filter is a non-blind approach for recovering a blurred image. As a result, there is a chance to reduce the additive noise in numerous locations. In comparison to the average and median filters, the Wiener filter is the best for removing noise.

2.3 Existing Algorithm

Super-Resolution Convolutional Neural Network (SRCNN). It is the first deep learning algorithm for image super-resolution SRCNN. It has a simpler architecture than VDSR. The SRCNN algorithm might perform well on some datasets and applications but it may struggle to capture fine specifications. Very Deep Super Resolution (VDSR) is intended to capture more complicated features and details because it is supposed to be very deep. The use of skip connections in residual learning can aid in the preservation and enhancement of high-frequency components. In several circumstances, VDSR has shown cutting-edge performance. PSNR evaluates image quality by comparing it to a reference image and taking into account the peak signal-to-noise ratio. In general, higher PSNR values suggest greater image quality. SSIM is a statistic that takes into account structural information in a picture. Greater structural similarity between the original and processed images is indicated by higher SSIM scores.

1. SRCNN

SRCNN stands for Super-Resolution Convolutional Neural Network. Chao Dong *et al.* [23], proposed a SRCNN has several encouraging characteristics. First, its structure is purposefully meant to be simple while still providing higher accuracy when compared to state-of-the-art example-based approaches. Second, with moderate training, the Super-Resolution Convolutional Neural Network (SRCNN) outperforms the bicubic baseline with only a few training iterations and outperforms the sparse-coding-based technique (SC). More training iterations may increase performance even further. For image super-resolution, a fully convolutional neural network is used. The network learns an end-to-end mapping between low and high-resolution images natively, with little pre/postprocessing beyond optimization. Make a connection between our deep learning-based SR approach and standard sparse-coding-based SR methods. This relationship directs the network structure's construction. Performing that deep learning can achieve good quality and speed in the classic computer vision challenge of super-resolution.

2. VDSR

Jiwon Kim *et al.* [24], an experimental work was done by this approach employs a very deep convolutional network based on the VGG network, which is utilized for ImageNet classification. It discovers that increasing network depth results in a significant gain in accuracy. Contextual information over vast image regions is efficiently exploited by cascading tiny filters many times in a deep network structure. However, with highly deep networks, convergence speed becomes a key issue during training. It generates a very simple but effective training approach. Learn entirely residuals and employ extremely high learning rates (more than one of SRCNN) afforded by configurable gradient cutting. This method performs better than SRCNN in terms of accuracy in the form of PSNR and SSIM.

3. PSNR (Peak Signal to Noise Ratio)

Quality is an essential criterion for all objects and their functions. Umme Sara *et al.* [25], is explained about PSNR is used to calculate the ratio of maximum feasible signal strength to distorting noise power, which influences the quality of its representation. This ratio of two images is calculated in decibels. Because signals have a relatively large dynamic range, the PSNR is commonly calculated as the logarithm term of the decibel scale.

The Peak signal-to-noise ratio is the most often used quality assessment metric for determining the quality of lossy image compression codec reconstruction. The signal is the actual data, while the noise is the error caused by compression or distortion. The PSNR is an approximation of the human impression of reconstruction quality in comparison to compression code. The PSNR value fluctuates from 30 to 50 dB for 8-bit data representation and from 60 to 80 dB for 16-bit data in image and video compression quality degradation. The acceptable range of quality loss in wireless transmission is around 20 - 25 dB.

PSNR is expressed as: psnr = 20*log10(255/rmse); ---(1)

4. SSIM

The structural similarity (SSIM) index is a grayscale image quality measuring metric. It computes the similarities between an image's luminance, contrast, and structure. The SSIM index is the result of multiplying these three factors. These highly interdependent pixels relate to additional crucial information about visual things in the image domain. Luminace masking refers to the process of making the distortion component of an image less noticeable at the image's edges. Contrast masking, on the other hand, refers to the process of making distortions in an image's texture less obvious. The perceived quality of photos and videos is estimated by SSIM. It compares the similarity between two images: the original and the recovered is experimented by Umme Sara *et al.* [25].

Structural Similarity Index Method can be expressed through these three terms as:

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \quad \text{---(2)}$$

Here, l is the luminance (used to compare the brightness of two images), c is the contrast (used to compare the ranges between the brightest and darkest regions of two images), and s is the structure (used to compare the local luminance pattern of two images to determine the similarity and dissimilarity of the images) α, β and γ are the positive constants.

$$l(x,y) = \frac{2\mu_x\mu_y+c1}{\mu^2_x+\mu^2_y+c1} \quad \text{---(3)}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y+c2}{\sigma^2_x+\sigma^2_y+c2} \quad \text{---(4)}$$

$$s(x,y) = \frac{\sigma_{xy}+c3}{\sigma_x+\sigma_y+c3} \quad \text{---(5)}$$

$$SSIM\ (x,y) = \frac{(2\mu_x\mu_y+c1)\ (2\sigma_x\sigma_y+c2)}{(\mu^2_x+\mu^2_y+c1)\ (\sigma^2_x+\sigma^2_y+c2)} \quad \text{---(6)}$$

3. Experimental Result

In an experimental investigation exploring the efficacy of preprocessing techniques in image processing, such as frame conversion, face detection and extracting the face, resizing an image, Grayscale image, noisy image, Median filter, and Wiener filter were applied to optimize the input data. Preprocessing had a considerable influence on the super-resolution performance of both SRCNN and VDSR, according to the experimental data. The use of preprocessing approaches improved model convergence reduced reconstruction errors, and improved overall image quality. The preprocessing techniques enhanced the SRCNN, which is known for being able to learn detailed features, by more accurately capturing and representing image information. Similarly, the VDSR model's deep architecture showed increased sensitivity to pre-processed inputs, resulting in superior super-resolution results. And using a combination of evaluations, such as PSNR and SSIM, may provide an enhanced evaluation of image quality.

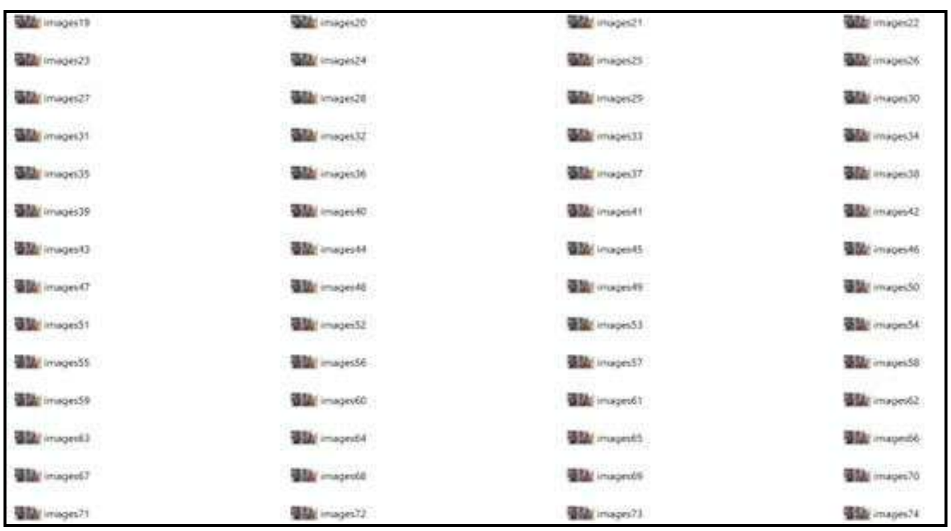

Figure 3: Frame Conversion

Figure 3 shows the Frame conversion of CCTV image, **Figure 4** shows the input image, **Figure 5** shows the Face detection of an input image and automatically chose the best image of a face detection and extract the face.

Figure 4: Input Image

Figure 5: Face Detection

Figure 6: Resize Image

Figure 7: Gray Scale Image

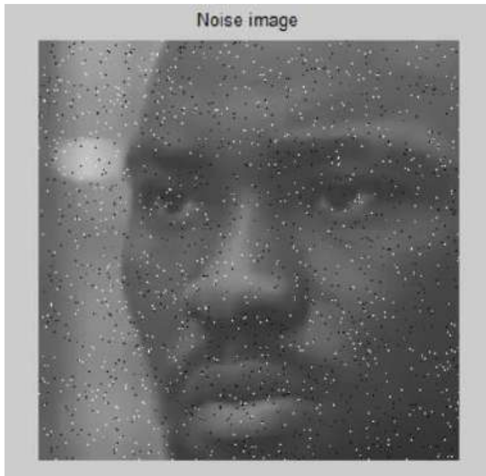

Figure 8: Noise Image

Figure 6 shows the resize of an image, Gray Scale and Noise Image is applied to the image shown in **Figure 7, Figure 8.**

Figure 9: Median filter image

Figure 10: Wiener Filter Image

Figure 9 shows a Median Filter and **Figure 10** shows a Weiner filter of an image.

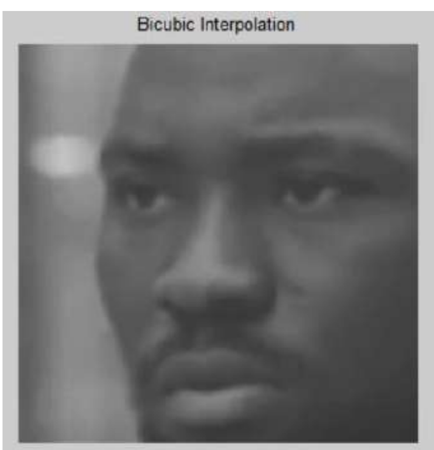

Figure 11: Bicubic Interpolation

Figure 12: SRCNN Reconstruction

After Preprocessing **Figure 11** shows the existing algorithm result of an image is Bicubic Interpolation and **Figure 12** shows the reconstructed image is SRCNN Reconstruction.

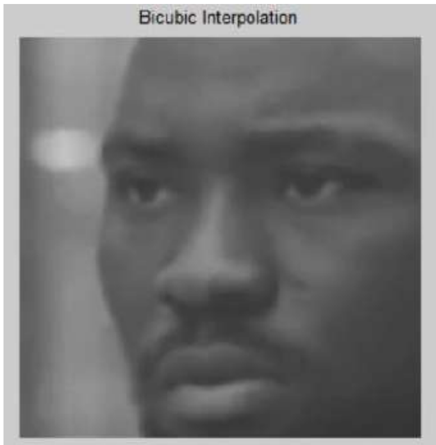

Figure 13: Bicubic Interpolation

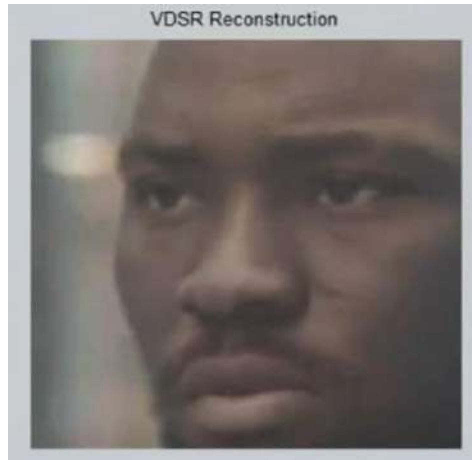

Figure 14: VDSR Reconstruction

Figure 13 shows the image it is Bicubic Interpolation and **Figure 14** shows the reconstructed image is VDSR Reconstruction.

4. RESULT AND DISCUSSION

This area gives a detailed perspective on the outcome that is gotten by preprocessing and existing algorithm which is acted in the working foundation of MATLAB. Here the conventional Neural Network methods for low resolution to high resolution. The test result and the presentation of the existing strategy is given in detail.

The experiments were carried out on Intel core i5 processor with 1.80 GHz, 8 GB RAM memory which works on windows 11 operating system. The computational time and the memory space may vary depending upon the system requirements; for this hardware specification the CCTV Videos dataset produces the results that are given below. The resultant dataset is with the existing algorithms to validate the efficiency and accuracy and in finding the best algorithm. PSNR and SSIM is used to evaluate quality of an image.

Table 1: Average PSNR/SSIM for scale factor ×2, ×3 and ×4 on Bicubic and SRCNN

Scale	Bicubic		SRCNN	
	PSNR	SSIM	PSNR	SSIM
*2	43.647558 dB	0.9968	43.906941dB	0.9970
*3	40.165135 dB	0.9918	40.253233 dB	0.9919
*4	37.601312 dB	0.9877	37.652520 dB	0.9878

Table 1 shows the result of Average value of scale factor *2,*3,*4 for PSNR and SSIM for Bicubic and SRCNN algorithm.

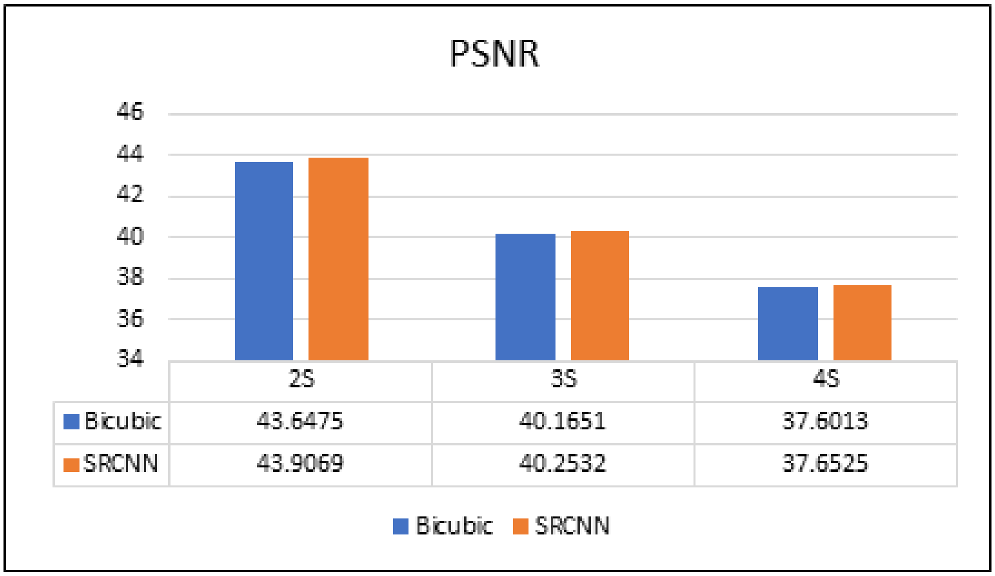

Figure 15: PSNR Value for Bicubic and SRCNN

Figure 15 shows the performance analysis of PSNR value for Bicubic and SRCNN algorithm

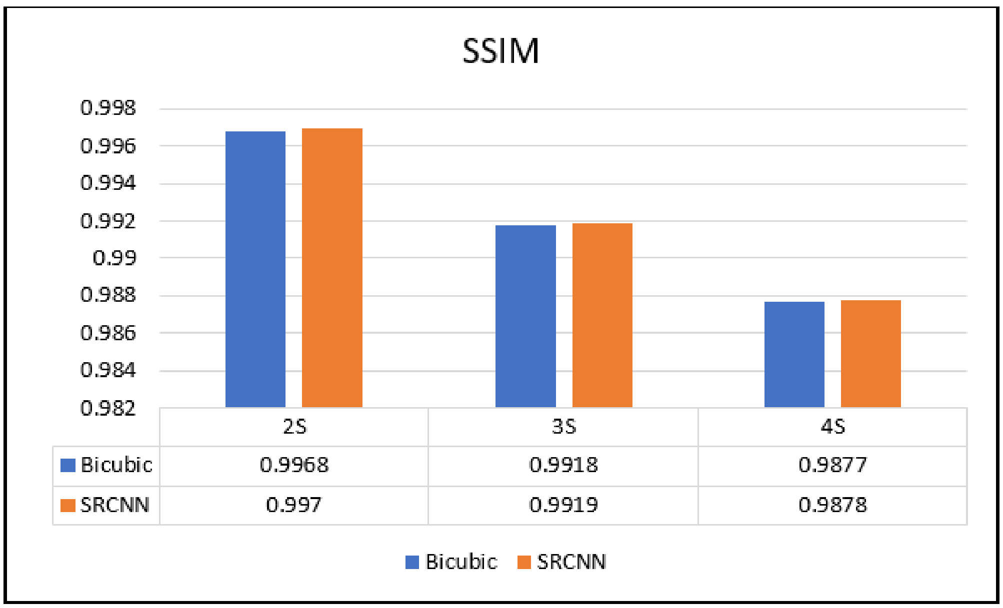

Figure 16: SSIM Value for Bicubic and SRCNN

Figure 16 shows the performance analysis of SSIM value for Bicubic and SRCNN algorithm.

Table 2: Average PSNR/SSIM for scale factor ×2, ×3 and ×4 on Bicubic and VDSR

Scale	Bicubic		VDSR	
	PSNR	SSIM	PSNR	SSIM
*2	52.937664 dB	0.9991	53.934817 dB	0.9995
*3	49.456887 dB	0.9965	49.553657 dB	0.9969
*4	47.040511 dB	0.9940	47.039371 dB	0.9942

Table 2 shows the result of Average value of scale factor *2,*3,*4 for PSNR and SSIM for Bicubic and VDSR algorithm.

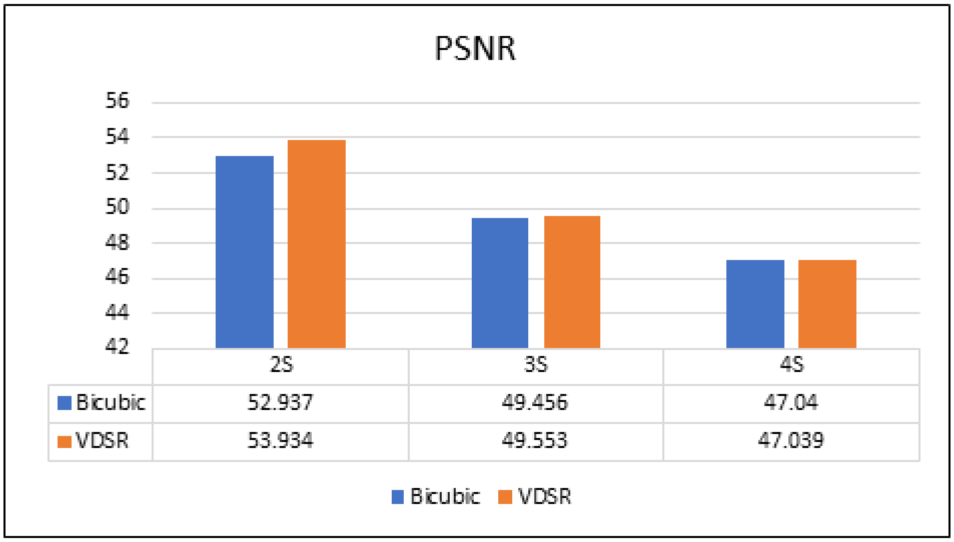

Figure 17: PSNR Value for Bicubic and VDSR

Figure 17 shows the performance analysis of PSNR value for Bicubic and VDSR algorithm

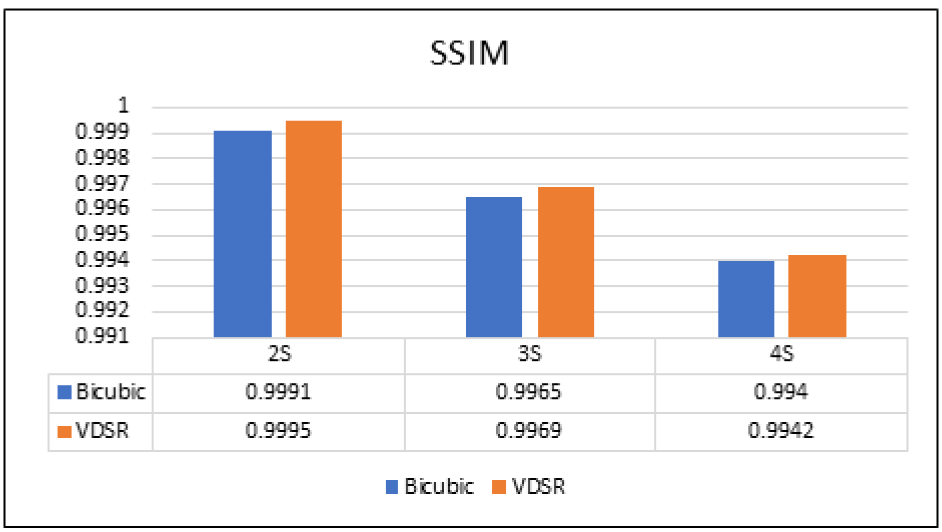

Figure 18: SSIM Value for Bicubic and VDSR

Figure 18 shows the performance analysis of SSIM value for Bicubic and VDSR algorithm.

For the above result, after analyzing the existing algorithm VDSR is better than SRCNN using SSIM and PSNR depends on the specific dataset, training parameters, and evaluation criteria. In general, SSIM assesses structural similarity between the original and produced images, whereas PSNR assesses reconstruction quality by comparing pixel values. SSIM is a better measure of imperceptibility in all aspects of an image rather than PSNR.

5. CONCLUSION

In this research work, Video processing in the context of digital imaging involves the application of various techniques to a series of images (frames) to analyze, improve or transform video data. Preliminary processing, SRCNN, and VDSR work together to provide a comprehensive and successful approach to image processing and super-resolution tasks. SRCNN provides a comprehensive framework for learning and mapping low- to high-resolution transformations, while VDSR takes it a step further by employing a very deep architecture to capture complex features. Together, the VDSR approach performs better than SRCNN methods in terms of accuracy, and visual improvements in results are immediately visible. And also, this approach is readily applicable to other image restoration problems such as denoising and compression artifact removal. It may conclude that SSIM maintaining structural details and perceptual quality is more important (e.g., image super-resolution, video compression), SSIM is a better measure of imperceptibility in all aspects of an image rather than PSNR.

6. REFERENCE

[1] Jaiswal, Varshali, Varsha Sharma, and Sunita Varma, "Comparative analysis of CCTV video image processing techniques and application: A survey", *IOSR Journal of Engineering* (IOSRJEN), Vol. 8, No.10, 2018, pp: 38-47.

[2] Rehmat Ullah, Hassan Hayat, Afsah Abid Siddiqui, Uzma Abid Siddiqui, Jebran Khan, Farman Ullah,Shoaib Hassan, Laiq Hasan,Waleed Albattah, Muhammad Islam,and Ghulam Mohammad Karami, "A real-time framework for human face detection and recognition in cctv images", *Mathematical Problems in Engineering* ,Vol. 2022, pp: 1-12.

[3] Kurniawan, Jason, Sensa GS Syahra, and Chandra K. Dewa, "Traffic congestion detection: learning from CCTV monitoring images using convolutional neural network", *Procedia computer science,* Vol.144, 2018, pp: 291-297.

[4] Bansod, Vishakha L, and Asha Ambhaikar, "Crowd Analysis System for Images of CCTV Camera", *International Journal of Recent Technology and Engineering* (IJRTE), Vol. 8, No.5, pp: 1113 – 1118, 2020.

[5] Vinay. A, Lokesh. A, Kamath V.R, Murty K.N.B, and Natarajan.S, "Enhancement of Degraded CCTV footage for forensic Analysis". *In International Conference on Innovative Computing and Communications*, Vol. 1, 2020, pp. 617-636,

[6] Smitha K, Akshitha G, Akhila B, Jagruthi E, "Deep Learning CNN for Accident Detection System from CCTV Videos", *Journal of Critical Reviews*, Vol.10, No. 01, 2022, pp: 149 – 158.

[7] Elhoseny, Mohamed, "Multi-object detection and tracking (MODT) machine learning model for real-time video surveillance systems", *Circuits, Systems, and Signal Processing,* Vol.39, No. 2, 2020, pp. 611-630.

[8] Bai, Yunfeng, Vladimir Berezovsky, Vasily Popov, "Digital Rock Core Images Super Resolution via SRCNN Based on Accelerated Bicubic Interpolation", ICNSER2020: *Proceedings of the 2nd International Conference on Industrial Control Network and System Engineering Research*, 2020, pp. 77–83.

[9] Ooi Yoong Khang, and Haidi Ibrahim, "Deep learning algorithms for single image super-resolution: a systematic review", *Electronics,* Vol. 10, No. 7, 2021, pp 1- 33.

[10] Mohan Amrita, Ramji Dwivedi, and Basant Kumar, "Image restoration of landslide photographs using SRCNN", *Recent Trends in Electronics and Communication*: Select Proceedings of VCAS 2020, vol 777, 2022, pp1249-1259.

[11] Elsaid, Nahla MH, and Yu-Chien Wu. "Super-resolution diffusion tensor imaging using SRCNN: a feasibility study". *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* Jul 2019, pp. 2830-2834.

[12] Zhangzong Zhao, Li Song, Rong Xie, Xiaokang Yang, "GPU Accelerated High-Quality Video/Image Super-Resolution", *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting* (BMSB), 2016, pp 1- 4.

[13] Vadim Romanuke, “An Improvement of the VDSR Network for Single Image Super-Resolution by Truncation and Adjustment of the Learning Rate Parameters”, *Applied Computer Systems,* vol. 24, No. 1, May 2019, pp. 61–68.

[14] Donghyeon Lee, Sangheon Lee, Ho Seong Lee, and Hyuk-Jae Lee, “Context-Preserving Filter Reorganization for VDSR-Based Super-resolution”, *IEEE International Conference on Artificial Intelligence Circuits and Systems* (AICAS), 2019.

[15] Reddy, M. Vasudeva, and T. Ramashri, "Underwater Image Enhancement Using Very Deep Super Resolution Technique", *I-Manager's Journal on Image Processing*, Vol 8, No. 2, 2021, pp. 9-14.

[16] Mehmet Cem Catalbas, “Modified VDSR-based single image super-resolution using naturalness image quality evaluator”, Signal, Image and Video Processing, Vol.16, No. 3, Aug 2022, pp. 661–668.

[17] Vint D, Di Caterina G, Soraghan J.J, Lamb R.A. and Humphreys D, “Evaluation of performance of VDSR super resolution on real and synthetic images”. *IEEE Sensor Signal Processing for Defence Conference* (SSPD), May 2019, pp. 1-5.

[18] Donghyeon Lee, Ho Seong Lee, Kyujoong Lee, Hyuk-Jae Lee, “Fast Very Deep Convolutional Neural Network with Deconvolution for Super-Resolution”, *Journal of Korea Multimedia Society*, Vol. 20, No. 11, Nov 2017, pp. 1750-1758.

[19] Chandra Shekhar Mithlesh, Dolley Shukla, Manisha Sharma, “Video to Image Conversion Techniques Video Frame Extraction”, *International Journal of Emerging Science and Engineering* (IJESE), Vol. 4, No. 3, Feb 2016, pp 30 -34.

[20] Alpika Gupta, Dr. Rajdev Tiwari, “Face Detection Using Modified Viola Jones Algorithm”, *International Journal of Recent Research in Mathematics Computer Science and Information Technology,* Oct 2014 – Mar 2015, Vol. 1, No. 2, pp.59-66.

[21] Rong Zhu, Yong Wang, “Application of Improved Median Filter on Image Processing”, *Journal of Computers*, Vol. 7, No. 4, Apr 2012, pp: 838 – 841.

[22] Iman Hussein AL-Qinani, “Deblurring Image and Removing Noise from Medical Images for Cancerous Diseases using a Wiener Filter”, *International Research Journal of Engineering and Technology* (IRJET), Vol. 04, No. 08, Aug -2017. pp 2354 – 2356.

[23] Dong, Chao, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks", *IEEE transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 2, Jun 2015, pp. 295-307.

[24] Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks". *In Proceedings of the IEEE conference on computer vision and pattern recognition*, Jun 2016, pp. 1646-1654.

[25] Umme Sara, Morium Akter, Mohammad Shorif Uddin, “Image Quality Assessment through FSIM, SSIM, MSE and PSNR—A Comparative Study”, *Journal of Computer and Communications*, Vol.7 No.3, Mar 2019, pp 8- 18.