Hybrid SVR-ACNN Model: Proposed Video Super-Resolution Method for Video Enhancement

[1]Padma Reddy A. M, [2]Udayarani
[1]Sai Vidya Institute of Technology, Rajanukunte
Via Yelahanka, Bengaluru, Karnataka 560064, India
Padmareddyam14@Gmail.Com

[2]Reva University
Kattigenahalli, Bengaluru
Karnataka 560064
India

ABSTRACT: *Video super-resolution techniques are the need of the hour for the high-resolution display devices as the current high-resolution videos are the basic question. Even though there are a large number of techniques employed for the video super-resolution, all these existing techniques face a hectic challenge at various conditions. Thus, this paper proposes an effective video resolution strategy using the hybrid Support vector regression- Actor Critic Neural Network (SVR-ACNN) model for video enhancement. The super-resolution images formed using the individual SVR model and Actor Critic Neural Network are integrated using the weighted average concept. The Actor Critic Neural Network is tuned optimally using the proposed Fractional-based Sine Cosine algorithm (F-SCA) that is responsible for the global optimal convergence. The experimentation of the proposed method utilizes three videos taken from the Cambridge-driving Labeled Video Database (CamVid), and the results are analyzed for three scaling factors. The outcome of the analysis proves that the proposed method offers a better super-resolution image with a better PSNR, SSIM, and SDME of 33.6447dB, 0.9398, and 45.2779, respectively.*

Subject Categories and Descriptors
[I.2.10 Vision and Scene Understanding]; Video: **[H.4.3 Communications Applications];** Videotex: [**H.5.1 Multimedia Information Systems];** Video

General Terms
Super-resolution, Video, Support Vector Regression, Neural Networks

1. Introduction

Digital video plays a major role in the day-to-day life, and videos [33] and images with High resolution are mainly employed as they probably minimize the computation cost required for display, processing, and analysis. For instance, one can say that the need for displaying the videos with Low Resolution (LR) videos namely, the Standard-Definition (SD) video signals on High Definition (HD) displays with greater quality. Most of the resolution issues present in the Surveillance videos are that they lose their resolution such that the required frame rate for the videos has to be guaranteed for dynamic scenes [10]. The issues of the resolution can be rectified using the suitable video enhancement strategies, and the main aim of the video enhancement is to unveil the information of the video that is found hid [13]. Video enhancement is approached as the formulation of the high resolution and high-quality video out of the low resolution video as per the requirements of the specific applications. Thus, video enhancement aims at improving the clarity of the input

video as most of the applications use the digital video for processing, verification of traffics, criminal issues, and so on that is required being with good clarity for analysis [2].

Video enhancement is performed based on the spatial domain and frequency domain. In the spatial domain, the pixels are manipulated directly in the space plain, and in the frequency domain, the spatial frequency spectrum of the image is modified [11]. Even though video enhancement derives high quality videos, there are some degradation factors affecting the quality enhancement process. The problems arrive due to the low contrast such that extraction of the object from the dark background is an issue. Moreover, the problem may be due to the expertise of the human operators and poor quality of the video [12]. The most common method employed for overcoming the risk factors of dealing with the surveillance videos includes the video Super-Resolution (SR) [1]. The video-based super-resolution aims at the generation of the high-resolution frames by complementing the details of the image pixels [2] [5]. The aim of the video SR is to generate the High-Resolution (HR) video frames using the sequence of Low-Resolution (LR) inputs. Video super-resolution drags a lot of attention in the last few years both on the academic side and on the industry [10] [11]. The available HR video devices are highly expensive to generate, store and transmit the HR videos and hence, the demand falls on the modern SR techniques such that the generation of the HR videos from LR ones is facilitated [3].

Video SR techniques enable us to perform the video coding/decoding [14], face video hallucination [15], video surveillance systems [16], remote sensing systems [17], medical image analysis [18] and stereoscopic video processing [19] [11] [4]. SR techniques are grouped into two main categories, such as single-image based and multi-frame based [20], [21]. Single-image based SR methods use interpolation based and example-based methods. Interpolation based methods exhibit low computational cost, but with limited restoration performance. Bayes theorem [34] and linear regression [39] have been utilized in video super-resolution. Researchers pay their interest on the example-based single-image SR, in which the external and/or internal exemplars are sheathed for learning the mappings using the low-resolution patch to yield the high-resolution patches. Example-based SR algorithms use nearest neighbor [22], neighbor embedding [23], sparse representation [24], anchored neighbor regression [25] or deep learning [26] [4]. A lot of existing multi-frame SR methods that model the long-term and short-term dependencies through the extraction of the subpixel motions in the video frames are developed in the literature. The fact of modeling is proven to be effective using the Recurrent Neural Networks (RNNs) [27] as they possess a higher degree of temporal dependency modeling in sequential data processing [7]. Video SR techniques can be used in conjunction with other image processing techniques, like medical image analysis [35, 36, 37, and 38].

This paper uses an effective video enhancement strategy using the proposed SVR-ACNN model. The LR video is transformed as the SR video and the model processes on the individual frames of the video. The quality of the video is enhanced, for which initially, the individual frames are sent to the SVR model and Actor Critic Neural Network (ACNN) individually and the obtained super-resolution image is subjected to the weighted average concept based on the weights. The obtained super-resolution image sustains the video quality and highlights the higher degree of contrast to make it effective for further video processing. ACNN is tuned using the proposed algorithm that is the integration of the fractional concept in the standard Sine Cosine Algorithm (SCA). The proposed F-SCA inherits the advantages of both the fractional theory and SCA that possesses better convergence and provides the global optimal weights to tune the ACNN.

The major contributions of the paper are:

F-SCA algorithm for tuning the optimal weights of ACNN: The weights of ACNN are tuned optimally using the proposed algorithm that is the integration of the fractional concept in the SCA. The global optimal weights are derived for ACNN that derives the superresolution video.

Hybrid SVR-ACNN model: The proposed hybrid SVR-ACNN model generates the superresolution image. Initially, the low resolution image is subjected to the high resolution individually using two models, SVR and ACNN, and finally, the high resolution image is formed using the weighted average.

The organization of the paper is:

Section 1 introduces the paper; section 2 elaborates the literature review of the superresolution images. Section 3 states the problem and the proposed method of super-resolution is discussed in section 4, section 5 details the results and discussion of the proposed method. Finally, section 6 concludes the paper.

2. Motivation

This section depicts a review of the literature on various existing video super-resolution methods. These research papers are taken and reviewed according to the recently published years based on the video super-resolution techniques.

2.1 Related Works
Armin Kappeler et al. [1] presented Convolutional Neural Networks (CNNs) that used the dimension of videos for increasing the spatial resolution. Consecutive frames were motion compensated and used as input to the CNN. Kun Li et al. [2] developed a method that automatically selected and obtained a super-resolved image. The results were based on the spatial-temporal characters, but the method was not effective. Wenhan Yang et al. [3] suggested Spatial-Temporal Recurrent Residual Networks (STR-

ResNet) for video superresolution. This network simultaneously models high-frequency details of single frames, the differences between HR and LR frames, as well as the changes of these adjacent detail frames. Dingyi Li and Zengfu Wang [4] designed a video SR algorithm that was able to handle large and complex motions adaptively. Amar B. Deshmukh and N. Usha Rani [5] designed a model that alleviated the resolution issues, and the designed model is termed as fractional-Grey Wolf optimizer-based kernel weighted regression model. The merit of the method is that there is no degradation in the quality of the image. The method provides a super-resolution image without degrading the quality of the LR image, but the performance is poor in case of greater value of Second Derivative like Measure of Enhancement (SDME). Yawei Li et al. [6] designed an adaptive factor that was integrated in the Non-Local Means (NLM) algorithms that overcame the drawback of fixed decaying factor and searching window. The robustness of the method was better, but it was dependent on the pixel-wise computation. Yan Huang et al. [7] designed a model, bidirectional recurrent convolutional network, for performing the multi-frame SR. The method offered low computational complexity and performed the order of magnitude that was faster when compared with other multi-frame SR methods, but this method was not applicable for large-scale high-resolution video. Di Chen et al. [8] modeled a method through the integration of the Compensationbased TV (CTV) regularization term with Multi-Non-local Low-Rank (MNLR) regularization term in the optimization algorithm. The algorithm minimized the negative impacts and possessed the capacity to withstand the noise, but the method suffered from noise effects while using a large database.

2.2 Challenges
• The process of generating the super-resolution frames may yield high-quality results, but the methods fail to take advantage of significant correlations between adjacent frames. Moreover, these methods possess high computational demands in case of a large number of frames [2].

• The performance of the Example-based super-resolution methods [4] is found to be better, but their results may be impossible and as they depend on the qualities of the training datasets.

• The super-resolution based on CNN model [1] suffers from high computational cost as they are time-consuming as well as it suffers from visual artifacts that occur due to the complex motions available in the video frames.

• Self-enhancement techniques [12] are straightforward for the implementation, but they suffered with the dark original videos as they lose the information at the time of pre-processing. The learning-based or example-based methods [2] possess highly complex computations as the result of the dictionary training and patch matching.

3. Problem Statement

The ultimate goal of the paper is to obtain the SR video using the LR videos and the conversion of the LR videos into the SR videos is based on the scaling factor such that the pixels in the image are modified without affecting the quality of the image. Consider an LR video V with n number of frames,

$$V = \{f_1, f_2, \ldots f_i, \ldots f_n\} \qquad (1)$$

where, V implies the LR video and f_i indicates the ith frame. The pixel location of the ith frame centered at the location, (g, h) is denoted as, $f_i(g, h)$. The dimension of the ith frame centered at (g, h) is given as, $(x \times y)$ and the scaling factor is denoted as, s. The LR image is converted to the HR image based on the scaling factor that is set by the user, and the HR image is derived using the SVR and ACNN model. The result from the two models is subjected to the weighted average to form the super-resolution image of the ith frame that is given as,

$$S = \{S_1, S_2, \ldots S_i, \ldots S_n\} \qquad (2)$$

The pixel location of the ith frame is denoted as (g, h), and the dimension of the ith super resolute frame is given as, $(s \times x, s \times y)$. The main intention of the paper is to transform the low resolution image of dimension (x, y) into the high resolution image of dimension, $(a \times b) \in [(s \times x) \times (s \times y)]$.

4. The Proposed Method of forming the Super-Resolution image using the Proposed SVR-ACNN Model

The aim of the proposed hybrid model is to form the super-resolution image, for which the hybrid model is developed using the SVR model and ACNN. Initially, the low resolution image is converted to the super-resolution image using the Support vector regression model and ACNN. The ACNN model is trained using the proposed F-SCA algorithm that is the integration of SCA and Fractional concept. The proposed algorithm tunes the network adaptively to generate the super-resolution image. The super-resolution image formed using the SVR model, and ACNN is averaged based on the weighted average to form the final output of the super-resolution image. The architecture of the proposed method of forming the super-resolution image is depicted in figure 1.

4.1 Super-Resolution Image using Support Regression Model
The LR image or in other words, the ith frame f_i of the video V is subjected to the SVR model and the first step in the SVR is the establishment of the kernel regression coefficient. The kernel regression coefficient depends not only on the location and density but also on the shape. The shape of the kernel regression is described as square, and its size is based on the user. Therefore, the kernel size is based on the scaling factor that is set by the user.
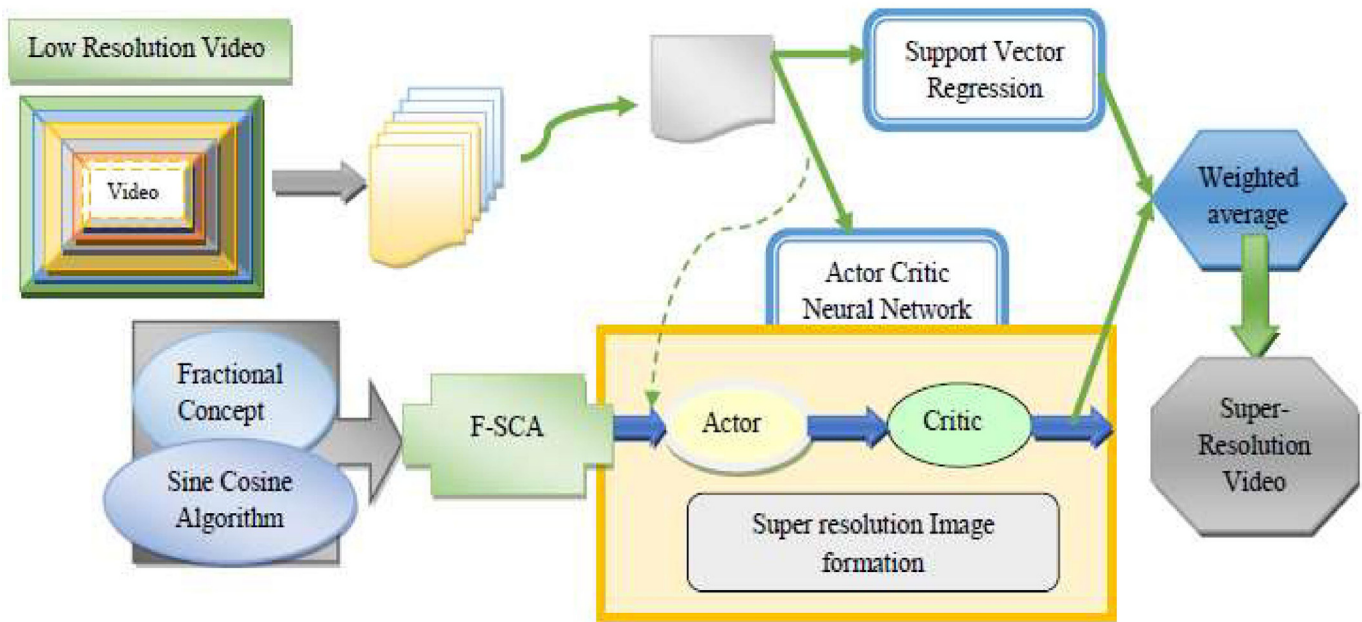
Figure 1. The architecture of developing the super-resolution image using the proposed hybrid model

The kernel regression coefficient is organized in the matrix format that is computed based on the distance value with respect to the center pixel. The kernel regression coefficient is denoted as,

$$C = \{K_{pq}; \qquad 0 \leq p; q \leq r \qquad (3)$$

The distance-based integer matrix is based on the kernel coefficient, and it is determined using the regression model [31].

4.1.1 Generation of the Super-resolution Image

The second step is the formation of the SR image that is obtained through the interpolation of the kernel regression matrix with the LR image. The pixel values of the newly generated SR image are generated using the support vector regression, and the dimension of the SR image is based on the scaling factor. The unknown pixel values of the SR image are calculated based on the neighboring pixels of the LR image. The SR image obtained using the SVR is denoted as, $f_1^i(g, h)$. The pixel of the SR image is determined as,

$$f_1^i(g, h) = \frac{1}{g \times h} \times \sum_{c_1=1}^{g} \sum_{c_2=1}^{h} f^i(g, h) \times R(g, h) \qquad (4)$$

where, c_1 and c_2 are the rows and columns of the sub-image or the ith frame, $(g \times h)$ represents the dimension of the frame, and $R(g, h)$ denotes the kernel regression matrix. $f^i(g, h)$ is the ith frame centered at (g, h). The kernel regression matrix is the dot product of the arbitrary that may or may not be known. The arbitraries refer to the pixels, and the unknown pixels are based on the similarity metric that is determined based on the neighboring pixels [31]. The dimension $(g \times h) \in [(s \times x) \times (s \times y)]$ is the dimension of the SR image.

4.2 Generation of the Super-resolution Image using the Proposed F-SCA-based Actor Critic Neural Network

The aim of the Actor-Critic Neural Network [30] is to generate the SR image using the ith frame of the LR video. The individual frames from the LR video are given as input to the ACNN that transforms the LR image to the HR image without any degradation in the quality of the image. ACNN consists of two modules, namely actor and critic modules that form the SR image. The advantage of ACNN is that it is adaptive and ACNN is a learning-based approach for forming the SR image. Each of the modules in ACNN is developed using the input layer, hidden layer, and the output layer and the actor module predicts the output that forms the input to the critic module. ACNN is adaptive due to the adaptive weights computed using the proposed F-SCA. The input to the actor module is the ith frame that is set with the scaling factor s multiplied with the weights of the input layer to present the input to the hidden layer, and output is predicted as the output of the actor layer. The output of the actor layer along with the SR image is presented to the input layer of the critic module that performs the successive computation in the hidden and the output layers to derive the SR image. In other words, the classifier obtains the pixel values of the SR image that is transformed based on the scaling factor. Figure 2 shows the architecture of Actor-Critic Neural Network employed to generate the SR images. The ACNN inherits reinforcement learning capabilities.

4.2.1 Actor Module

Actor module is the main module of ACNN, and the actor employs a parameterized method, named as a Neural Network (NN), and the NN with the single hidden layer serves as an actor module. The input to the actor module is I that carries the information of the test statistics and
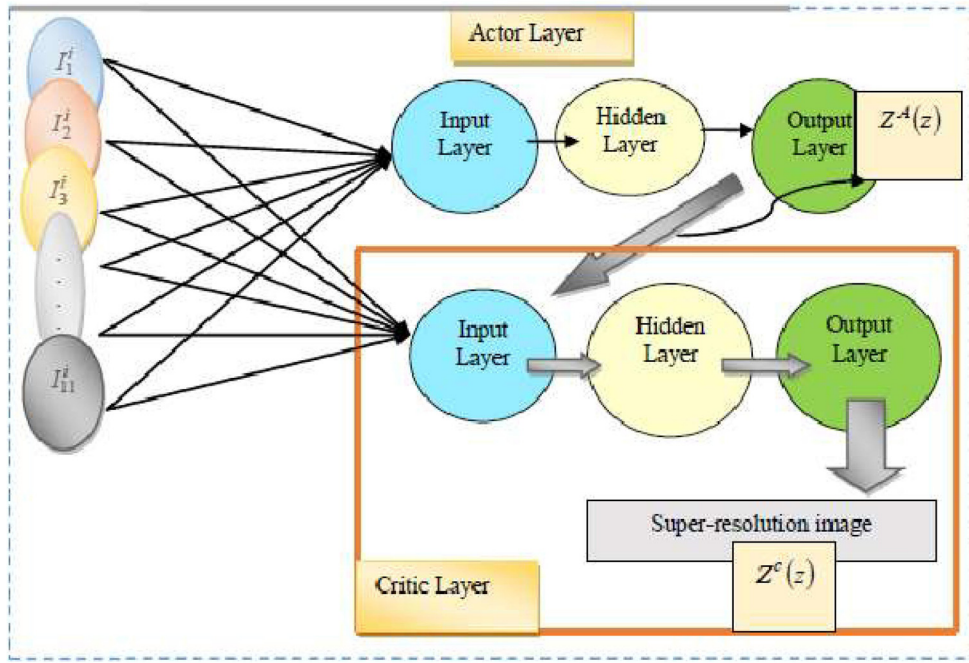
Figure 2. The architecture of Actor-Critic Neural Network for SR image formation

the Eigenvalues of the channel model. Thus, the successive computations engaged in the input and the hidden layers contribute to the output that forms the input to the critic layer. Thus, the output from the actor layer is represented as,

$$O^A(t) = w^A(t) \times \sigma(u^A(t) \times f_i^A(x, y)) \qquad (5)$$

where, $u^A(t)$ is the weight between the hidden and the output layers, $w^A(t)$ denotes the weight between the input and the hidden layers. The input vector that is the *ith* video frame, denoted as, $f_i^A(x, y)$, forms the input to the actor module. The output of the actor module is denoted as, $O^A(t)$. The dimension of the *ith* frame is converted from $(x \times y)$ to $(a \times b) \in [(s \times x) \times (s \times y)]$ using the scaling factor. This dimensionally converted *ith* frame f_i using the scaling factor is the actual input to the actor layer. Thus, the dimension of the input vector is given as, $(a \times b)$. The weights of the actor module are described as, $W_1 \in \{w^A(t),$ $u^A(t)\}$. The activation function depends on the hyperbolic tangent function and is denoted as,

$$\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \qquad (6)$$

The activation function is referred to as hyperbolic tangent function as is given in equation (6). The activation function is denoted as, $\sigma(z)$ for simplicity. $W_1 \in \{w^A(t), u^A(t)\}$ denotes the weight of the actor module.

4.2.2 Critic Module
The critic module is the module that generates the SR image, and the input to the critic module is the output from the actor module and the *ith* frame f_i of dimension

$(a \times b)$ that denotes the SR image. The critic layer determines the unknown pixel value of the frame. The adaptive weights are used to undergo a smooth prediction with better accuracy. The output from the critic layer is represented as,

$$O^B(t) = w^B(t) \times \sigma(u^B(t) \times f_i^B(x, y)) \qquad (7)$$

$$O^B(t) = w^B(t) \times \sigma^B(t) \qquad (8)$$

where, $f_i^B(a, b) = \lfloor f_i^A(a, b), O^A(t) \rfloor$ is the input to the critic module, $u^B(t)$ is the weight between the input and the hidden layers, and $w^B(t)$ is the weight between the hidden and the output layers. The weights of the critic layer are denoted as, $W_2 \in \{w^B(t), u^B(t)\}$. The output from the critic layer generates the unknown pixel values of the frame that denote the SR image, and the output of the critic layer is denoted as, $O^B(t)$.

4.2.3 Proposed F-SCA for tuning the Optimal Weights of ACNN
The main role of the proposed algorithm is to generate the optimal weights that change adaptively such that the SR image is generated effectively. The proposed F-SCA algorithm is the integration of SCA [28] and the fractional concept [32] such that the proposed F-SCA inherits the advantages of both the SCA and the fraction concept. The fractional concept possesses the tendency to avoid the convergence to the local optimum and causes the convergence to the global optimal solution. The fractional theory in the proposed algorithm increases its convergence rate and improves the performance of the SCA. Moreover, the fractional theory keeps a record of the past events and hence, the proposed F-SCA holds

the inherent memory property. The proposed F-SCA algorithm converges fast to the global optimal solution and provides a better optimization experience.

SCA [28] is a population-based optimization algorithm, and they begin the optimization with the generation of the random solutions that are verified using the objective function. The performance of the algorithm is enhanced using a set of rules that progress the optimization algorithm. The population-based algorithms look for the optimal solution and hence, the solution cannot be generated in a single iteration. The generation of random solutions and the iteration contribute to the chances of converging to the global optimal solutions. The algorithm consists of two phases, such as the exploration phase and the exploitation phase. In the exploration phase, the random solutions are generated, and they change continuously based on the objective functions in the successive iterations. The degree of randomness is high in case of the exploration phase such that the selection of the prominent areas in the search space requires the higher degree of the randomness. The random process is less in the exploration phase. Therefore, with the random set of solutions, the probability of determining the global optimal solution increases. The advantage of the SCA is that it is capable of solving the real optimization problems with the unknown search spaces and the algorithm uses the sine and cosine functions for exploring and exploiting the solution between the search spaces with the aim of converging to the best solution. The algorithmic steps are as follows:

Initialization: The weights of ACNN are initialized randomly, and they are iterated for the global optimal weights that ensure the effectiveness of the ACNN classifier. The initialization process follows the initialization of SCA, wherein the solutions are represented as, $z_{\hbar}; 1 \leq \hbar \leq \Re$, where $\Re$ is the size of the population.

Evaluate the Objective Function: The objective function aims at the minimization problem that is based on the error of the network. The objective function of all the search agents is updated, and the search agent with the minimum value of the objective function is chosen as the best search agent, and the position is updated. The search agent referred corresponds to the weights of ACNN.

$$\rho = \sum O^B(t) - O^{Ground}(t) \qquad (9)$$

where, ρ denotes the objective function. The output of ACNN and the ground truth is given as, $O^B(t)$ and $O^{Ground}(t)$, respectively.

Update the parameters: The main parameters in SCA are χ_1, χ_2, χ_3 and χ_4, and the parameter χ_1 symbolizes the movement direction or in other words, it indicates the next position regions and the direction may be either between the source and the destination or outer. The parameter χ_2 defines if the movement is towards or away from the des-

tination. The third random parameter χ_3 defines the random weights for the destination to stochastically emphasize and deemphasize the desalination impacts that define the distance. The last random parameter, χ_4 switches between the sine and the cosine components. The name SCA is due to the switching between the sine and the cosine components.

Update the Position using the proposed F-SCA: The proposed algorithm to update the position of the search agent is the modification of the SCA with the fractional concept. The position update follows two conditions: The two conditions are with respect to the fourth random parameter χ_4. The position of the search agent when the random number lies below the value 0.5 is given as,

$$Z(t + 1) = Z(t) + \chi_1 \times sin(\chi_2) \times |\chi_3 . M^t - Z(t)| \qquad (10)$$

where, $Z(t + 1)$ denotes the position of the search agent in the $(t + 1)^{th}$ iteration, χ_1, χ_2, χ_3 and χ_4 are the random numbers. The position of the destination search agent in denoted as, M^t and || stands for the absolute value. The random number χ_4 varies in the range 0 and 1. The standard equation of SCA is modified with the fractional theory, and the equation modification is inherited below. The fractional terms used in the equation use the history for the position update of the search agent at any iteration, and it enhances the performance optimization process. Moreover, the convergence of the algorithm is improved, and the global optimal solution is derived. The equation (10) is rewritten as,

$$Z(t + 1) - Z(t) = \chi_1 \times sin(\chi_2) \times \chi_3 . M^t - \chi_1 \times sin(\chi_2) \times Z(t)| \qquad (11)$$

$$D^{\beta}[Z(t + 1)] = \chi_1 \times sin(\chi_2) \times \chi_3 . M^t - \chi_1 \times sin(\chi_2) \times Z(t) \qquad (12)$$

where, $D^{\beta}[Z(t + 1)]$ indicates the differential component and it is the difference between the position of the search agent in the current and the previous iterations. Based on the fractional concept, the fractional theory is given as,

$$Z(t + 1) - \beta . Z(t) - \frac{1}{2} \times \beta \times Z(t - 1) - \frac{1}{6} \times (1-\beta) \times Z(t - 2) -$$

$$\frac{1}{24} \times \beta \times (1-\beta) \times (2-\beta) \times Z(t - 3) =$$

$$\chi_1 \times sin(\chi_2) \times \chi_3 . M^t - \chi_1 \times sin(\chi_2) \times Z(t) \qquad (13)$$

where, β is the fractional coefficient constant. Thus, the position of the search agent is given as,

$$Z(t + 1) = \beta . Z(t) + \frac{1}{2} \times \beta \times Z(t - 1) + \frac{1}{6} \times (1-\beta) \times Z(t - 2) -$$

$$\frac{1}{24} \times \beta \times (1-\beta) \times (2-\beta) \times Z(t - 3) +$$

$$\chi_1 \times sin(\chi_2) \times \chi_3 . M^t - \chi_1 \times sin(\chi_2) \times Z(t) \qquad (14)$$

$$Z(t+1)=Z(t)+[\beta-\chi_1\times sin\,(\chi_2)]+\frac{1}{2}\times\beta\times Z(t-1)+$$

$$\frac{1}{6}\times(1-\beta)\times Z(t-2)+\frac{1}{24}\times\beta\times(1-\beta)\times(2-\beta)\times Z(t-3)+$$

$$\chi_1\times sin\,(\chi_2)\times\chi_3\,.\,M^t \qquad (15)$$

The position of the search agent is updated based on the position of the search agent at time t, fractional coefficient β, and the position of the search agents in the previous iterations. The above equation is employed for updating the position of the search agent when the position of the destination is better than the position of the search agent, i.e., when $M^t > Z(t)$. The cosine component of the standard SCA is given as,

$$Z(t+1)=Z(t)+\chi_1\times cos\,(\chi_2)\times|\chi_3\,.\,M^t-Z(t)| \qquad (16)$$

The equation (16) is rewritten as,

$$Z(t+1)-Z(t)=\chi_1\times cos\,(\chi_2)\times\chi_3\,.\,M^t-$$

$$\chi_1\times cos\,(\chi_2)\times Z(t) \qquad (17)$$

$$D^\beta[Z(t+1)]=\chi_1\times cos\,(\chi_2)\times\chi_3.M^t-\chi_1\times cos\,(\chi_2)\times Z(t) \quad(18)$$

The integration of the fractional concept in the above equation is given as,

$$Z(t+1)-\beta.Z(t)-\frac{1}{2}\times\beta\times Z(t-1)-\frac{1}{6}\times(1-\beta)\times Z(t-2)-$$

$$\frac{1}{24}\times\beta\times(1-\beta)\times(2-\beta)\times Z(t-3)=$$

$$\chi_1\times cos\,(\chi_2)\times\chi_3\,.\,M^t-\chi_1\times cos\,(\chi_2)\times Z(t) \qquad (19)$$

$$Z(t+1)=\beta.Z(t)+\frac{1}{2}\times\beta\times Z(t-1)+\frac{1}{6}\times(1-\beta)\times Z(t-2)-$$

$$\frac{1}{24}\times\beta\times(1-\beta)\times(2-\beta)\times Z(t-3)+$$

$$\chi_1\times cos\,(\chi_2)\times\chi_3\,.\,M^t-\chi_1\times cos\,(\chi_2)\times Z(t) \qquad (20)$$

$$Z(t+1)=Z(t)+[\beta-\chi_1\times cos\,(\chi_2)]+\frac{1}{2}\times\beta\times Z(t-1)+$$

$$\frac{1}{6}\times(1-\beta)\times Z(t-2)+\frac{1}{24}\times\beta\times(1-\beta)\times(2-\beta)\times Z(t-3)+$$

$$\chi_1\times cos\,(\chi_2)\times\chi_3\,.\,M^t \qquad (21)$$

The position update equations can be employed for the search spaces with larger dimensions and the cyclic transition between the sine and cosine function allows the convergence to the best solution and enables the exploitation of the solution between the two solutions. The search space is even discovered out of the space between the destinations at the time of exploration. Thus, a proper balance between the exploration and the exploitation phases of the search space is obtained such that the global optimal solution is derived. The random number χ_1 is based on the current iteration and the maximum number of the iterations. The search space exploitation is defined in the interval [-1, 1].

The optimization process is progressed with the generation of the random solution and processes the solutions for exploring the best solution. The best solution determined is derived and is saved as the destination point such that the solutions of the successive iterations are updated based on the destination point. Moreover, the sine and the cosine range is updated with the increase in the iteration number, and the process is repeated for the

Proposed F-SCA	
1	**Input: Population of the search agent**
2	**Output: Global optimal solution (Best position of the search agent)**
3	**Begin**
4	Population Initialization
5	Do
6	Evaluate the search agents based on the objective function
7	Update the solutions
8	Update the random parameters χ_1,χ_2,χ_3, and χ_4
9	Update the position of the search agents based on equation (21).
10	While $t < \max\ number\ of\ iterations$
11	Return the best solution obtained till now as the global optimal solution
12	Terminate

Figure 3. Pseudo code of the proposed F-SCA

maximum number of the iterations. The generation of the global optimal solution is due to the following reasons: The random solutions generated are improved with the increase in the iteration count and the various search spaces are exploited when the sine and cosine spaces exceed 1 and -1 that allows the adaptive transition. The SR image obtained using ACNN is notated as, $f_2^i(g, h)$. Figure 3 shows the pseudo code of the proposed algorithm. At first, the weights of ACNN and the population of SCA are initialized. Then, the objective function is calculated for every search agent and the SCA parameters are updated. Afterthat, the position of the search agents are updated using equation (21). This process is repeated until the maximum iteration reaches. Finally, the best solution is determined.

4.3 Generation of the SR image using the Hybrid SVR-ACNN Model

The SR images obtained using both the SVR model and ACNN are subjected to the weighted average to form the SR image. The video enhanced output using the proposed hybrid SVRACNN model assures the SR video with good quality, and it highlights that there is no degradation in the quality of the video. The SR image obtained using the proposed model is given as,

$$S^i(g, h) = \beta \times f_1^i(g, h) + (1 - \beta) \times f_1^i(g, h) \qquad (22)$$

where, β represents the constant and $S^i(g, h)$ is the SR image. The dimension of the SR image $S^i(g, h)$ is $(a \times b) \in [(x \times s) \times (y \times s)]$.

5. Results and Discussion

The section deliberates the results and discussion of the proposed method of video resolution along with the comparative analysis in order to prove the effectiveness of the proposed method.

5.1 Experimental Setup

The software tool used for the implementation of the proposed technique is MATLAB that operates in the Windows 8 operating system.

5.2 Dataset Description

The database employed for the experimentation is taken from [29]. The Cambridge-driving Labeled Video Database (CamVid) [29] is the first collection of videos with object class semantic labels, complete with metadata. The database provides ground truth labels that associate each pixel with one of 32 semantic classes.

5.3 Experimental Analysis

The experimental analysis of the proposed method is highlighted in this section. Figure 4, figure 5, figure 6, and figure 7 shows the sample results of the experiment. Three videos, video 1, video 2, and video 3 and three frames each from those videos are considered for the analysis. The input frames are depicted in figure 4 a), figure 4 b), and figure 4 c). The HR images obtained for the scaling factors 2, 3, and 4 are depicted in figure 5, figure 6, and figure 7, respectively. The HR images obtained using the proposed method for the frame 1, frame 2, and frame 3, respectively, for the scaling factor 2 are given in figure 5 a), figure 5 b), and figure 5 c), respectively. The HR images obtained using the proposed method for the three frames with the scaling factor 3 are given in figure 6 a), figure 6 b), and figure 6 c), respectively. Similarly, the HR images obtained for the three frames with the scaling factor 4 are given in figure 7 a), figure 7 b), and figure 7 c), respectively.

5.4 Performance Metrics

The performance metrics employed for the analysis include the following:

i) Peak Signal-to-Noise Ratio (PSNR): PSNR is the measure that determines the quality of the image and the better method assures the maximum value of PSNR. PSNR is represented in decibel (dB).

ii) Second Derivative like Measure of Enhancement (SDME): The SDME metric identifies the visual quality of the enhanced image by calculating the second order measure. The expression for the SDME metric is
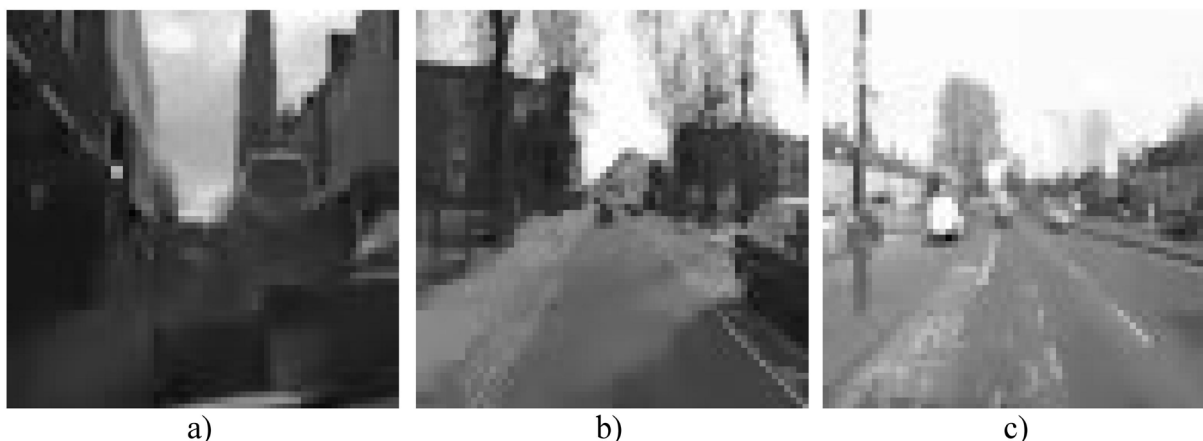

a) b) c)

Figure 4. Sample results of the experiment a) LR image from Video 1 b) LR image from Video 2 c) LR image from Video 3
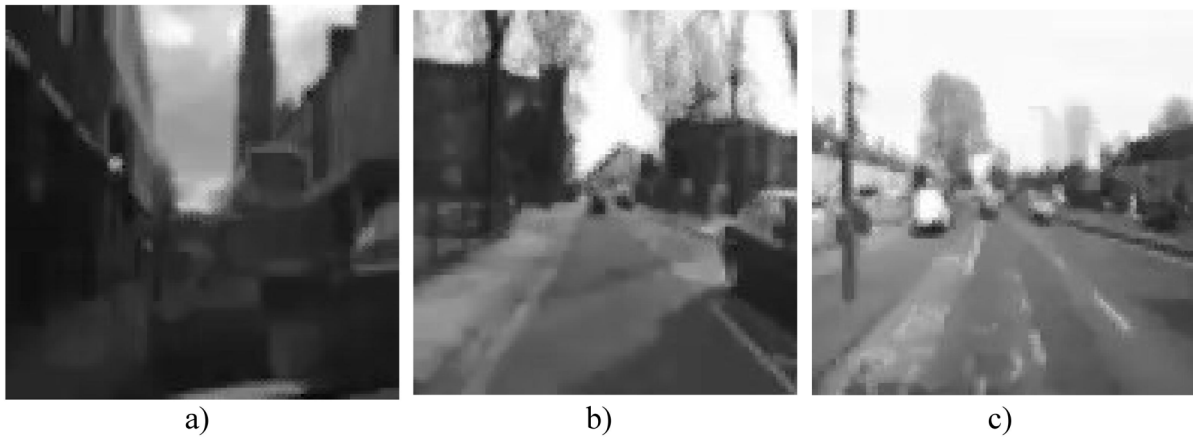
Figure 5. HR images using the scaling factor 2 a) for video 1 b) for video 2 c) for video 3
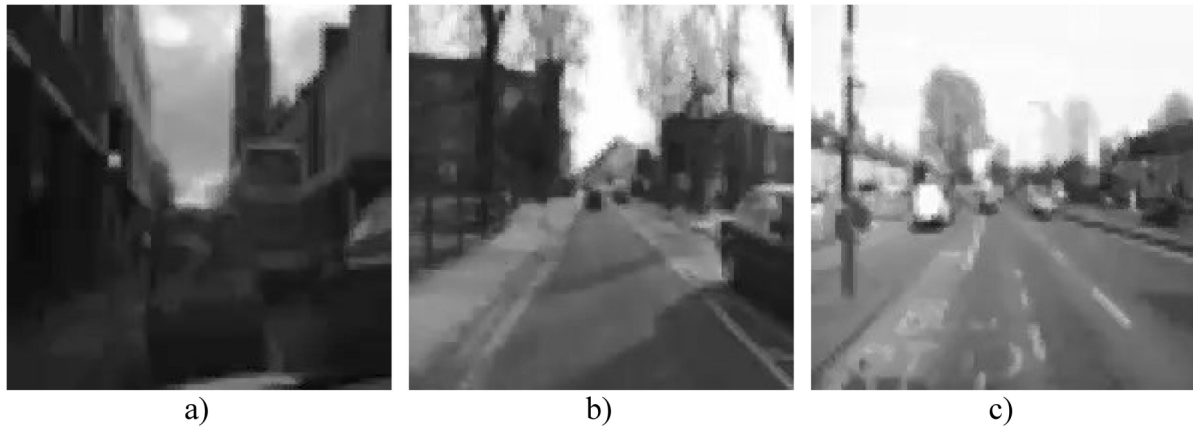

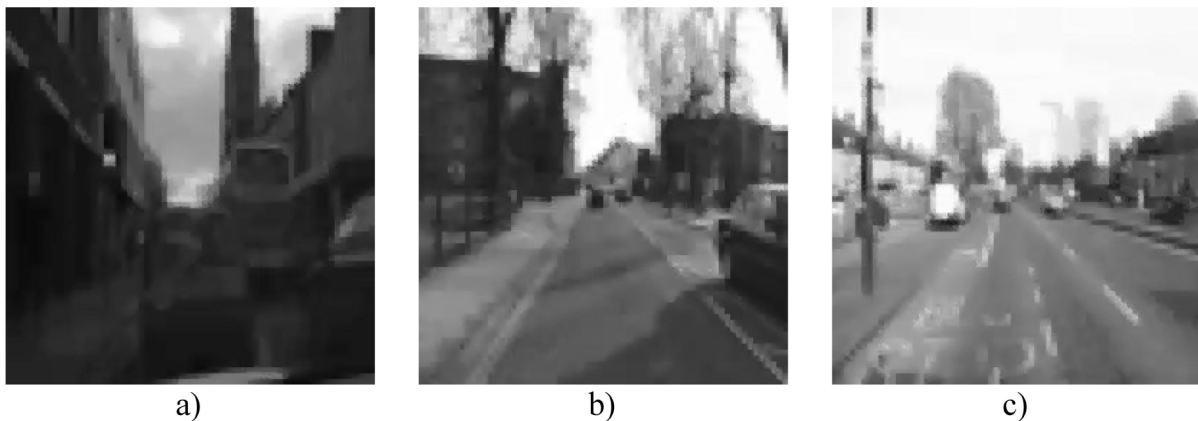Figure 6. HR images using the scaling factor 3 a) for video 1 b) for video 2 c) for video 3

Figure 7. HR images using the scaling factor 4 a) for video 1 b) for video 2 c) for video 3

expressed as follows,

$$SDME = - \frac{1}{v_1 \times v_2} \sum_{i=1}^{v_1} \sum_{j=1}^{v_2} 20 \, ln \left| \frac{K_{max}^{i,j} - 2K_{cen}^{i,j} + K_{min}^{i,j}}{K_{max}^{i,j} + 2K_{cen}^{i,j} + K_{min}^{i,j}} \right| \quad (23)$$

where, $K_{max}^{i,j}$, $K_{min}^{i,j}$, $K_{cen}^{i,j}$ refer to the maximum pixel value, minimum pixel value, and the center pixel value, respectively, v_1 and v_2 indicate the image blocks.

iii) Structural Similarity (SSIM) Index: SSIM is a measure for predicting the perceived quality of the SR image and the effective method assures the maximum value of SSIM.

5.5 Performance Analysis
This subsection shows the performance analysis of the proposed method by varying the hidden neurons.

5.5.1 Performance Analysis based on PSNR

Figure 8 shows the performance analysis of the proposed method based on PSNR for video 1, video 2, and video 3. Figure 8 (a) shows the PSNR of the proposed method for video 1. For the scaling factor 4, the proposed Hybrid SVR-ACNN with neurons 5 has the PSNR of 31.69dB. The proposed classifier has the PSNR of 31.86dB, 31.85dB, and 31.96dB when the number of neurons is 10, 15, and 20, respectively, for the scaling factor 4. The PSNR of the proposed method using video 2 is provided in figure 8 (b). The proposed classifier has the PSNR of 31.44dB, 31.65dB, 31.44dB, and 31.48dB when the number of hidden neurons is 5, 10, 15, and 20, respectively, for the scaling factor 4. Similarly, the PSNR curve of the proposed classifier for video 3 is provided in figure 8 (c). For the scaling factor 3, the proposed classifier has the PSNR of 31.36dB, 31.08dB, 31.55dB, and 31.54dB, when the number of hidden neurons is 5, 10, 15, and 20.

5.5.2 Performance Analysis based on SDME

Figure 9 shows the performance analysis of the proposed method based on SDME for video 1, video 2, and video 3. Figure 9 (a) shows the SDME of the proposed method for video 1. For the scaling factor 2, the proposed Hybrid SVR-ACNN has the SDME of 50.67dB, 50.68dB, 51.03dB, and 50.55dB, respectively, when the number of hidden neurons is 5, 10, 15, and 20. Figure 9 (b) depicts the SDME curve of the proposed classifier for video 2. The proposed classifier has the SDME of 46.54dB, 46.78dB, 46.69dB, and 46.77dB when the number of hidden neurons is 5, 10, 15, and 20, respectively, for the scaling factor 3. The SDME curve of the proposed classifier for video 3 is provided in figure 9 (c). For the scaling factor 4, the proposed classifier has the SDME of 54.59dB, 54.80dB, 54.56dB, and 54.92dB, when the number of hidden neurons is 5, 10, 15, and 20.
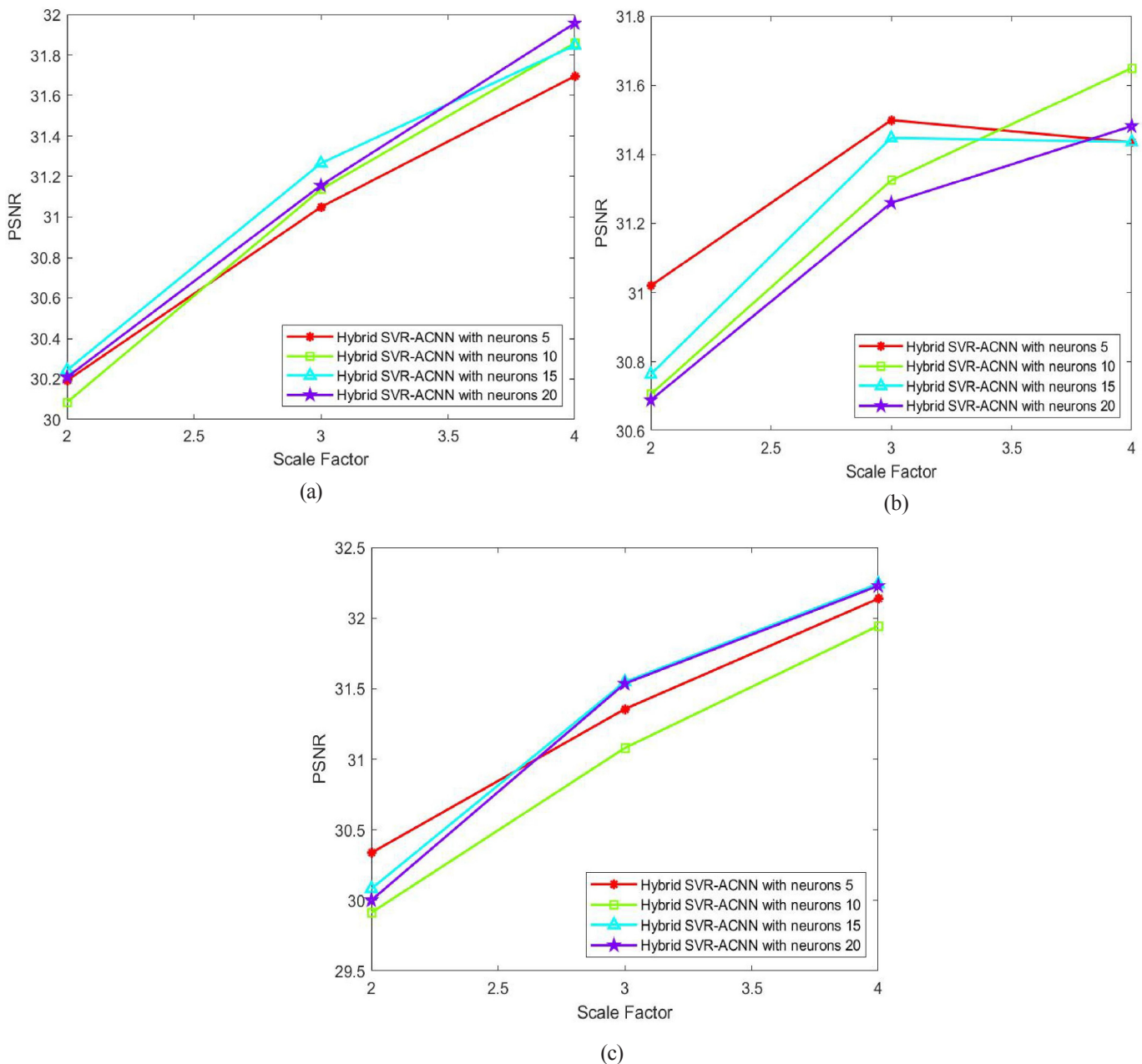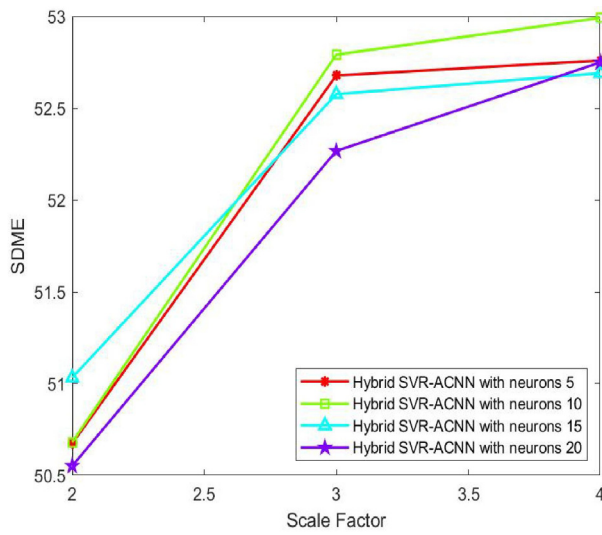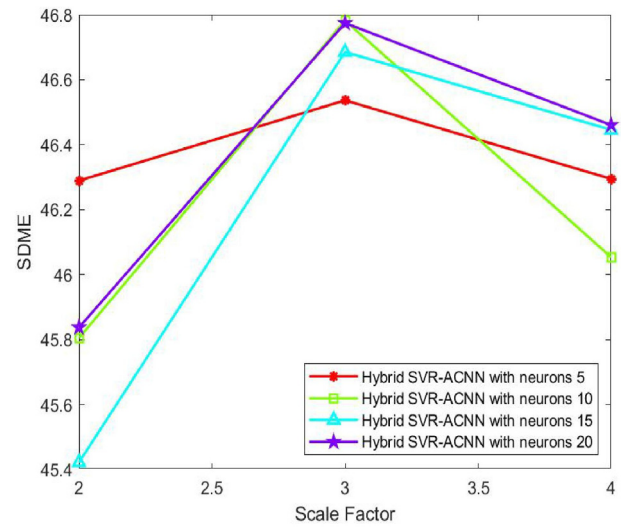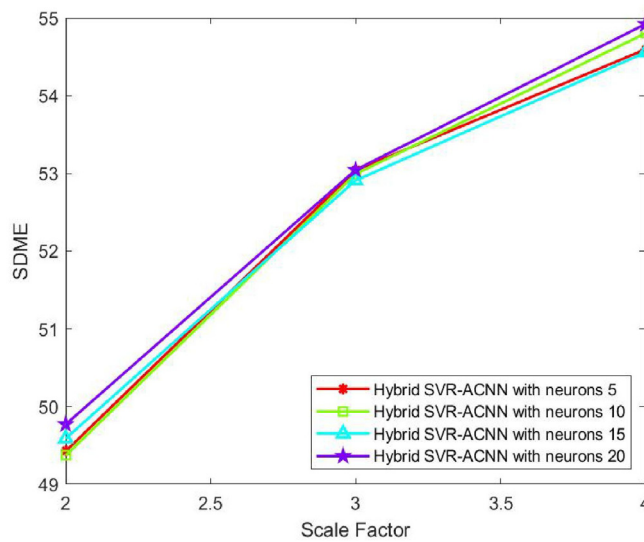

(a)

(b)

(c)

Figure 8. Performance analysis based on PSNR (a) video 1 (b) video 2 (c) video 3
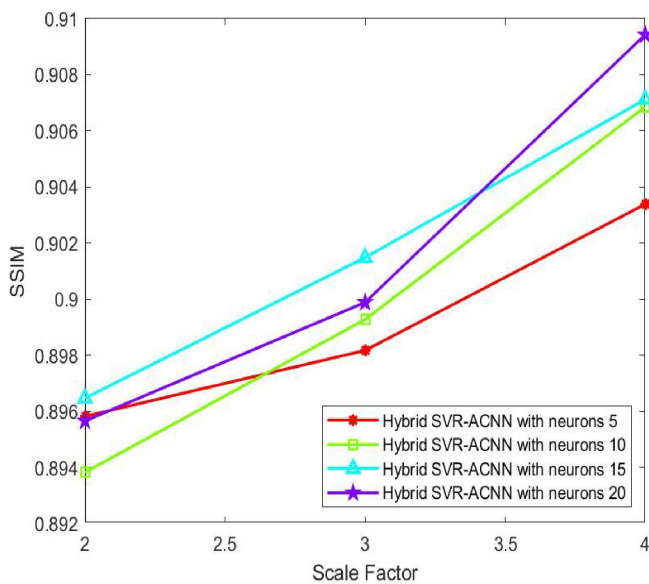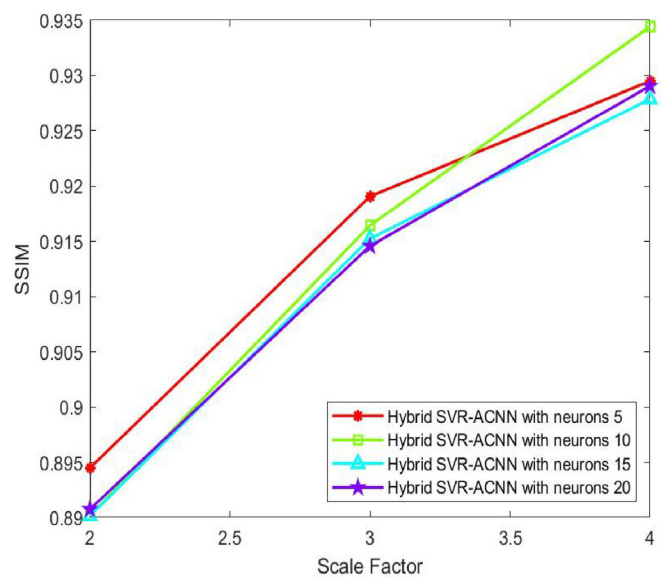
(a)

(b)

(c)

Figure 9. Performance analysis based on SDME (a) video 1 (b) video 2 (c) video 3
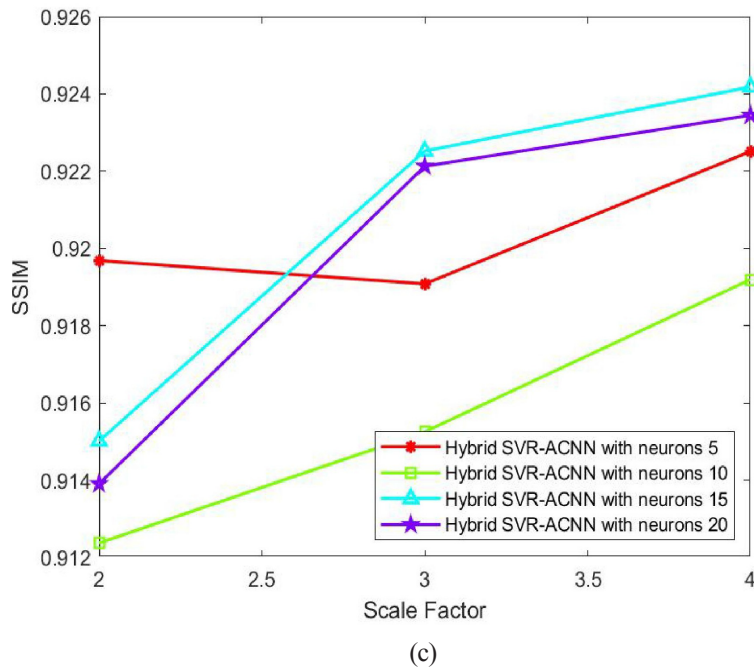
(a)

(b)

(c)

Figure 10. Performance analysis based on SSIM (a) video 1 (b) video 2 (c) video 3

5.5.3 Performance Analysis based on SSIM

The performance analysis of the proposed method based on SSIM for video 1, video 2, and video 3 is shown in Figure 10. Figure 10 (a) shows the SSIM curve of the proposed classifier for video 1. For the scaling factor 4, the proposed Hybrid SVR-ACNN classifier has the SSIM of 0.90 when the number of hidden neuron is 5, and the proposed classifier has the SSIM of 0.91, when the number of hidden neurons is 10, 15, and 20.

Figure 10 (b) depicts the SSIM curve of the proposed classifier for video 2. For the scaling factor 3, the proposed classifier has the SSIM of 0.92 when the number of hidden neurons is 5, 10, and 15. When the number of hidden neurons is 20, the proposed classifier has the SSIM of 0.91 for the scaling factor 3. The SSIM curve of the proposed classifier for video 3 is provided in figure 10 (c). For the scaling factor 4, the proposed classifier has the SSIM of 0.92, when the number of hidden neurons is 5, 10, 15, and 20.

5.6 Competing Methods

The methods employed for the analysis include ACNN [30], Recurrent Residual Network (RNN) [3], Convolutional Neural Network (CNN) [1], KNN, and Patch-based denoising [9]. The performance of these methods is compared with that of the proposed method in order to prove the effectiveness of the proposed method.
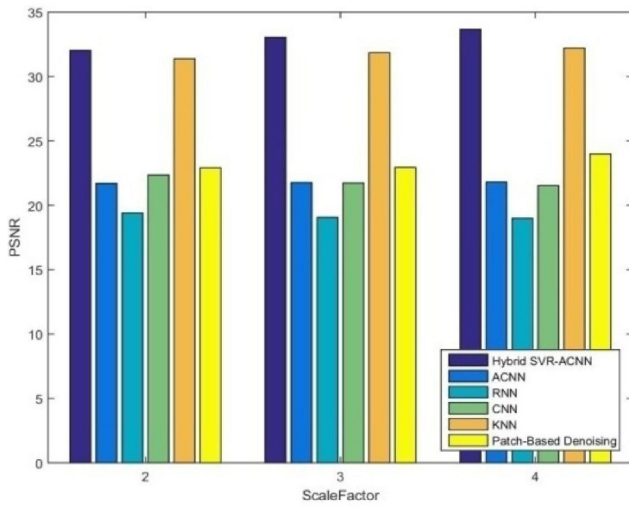
5.7 Comparative Analysis

This section presents the comparative discussion of the video SR methods based on the three evaluation metrics.

PSNR: The analysis based on the PSNR of the methods is progressed using three videos that are depicted in figure 11. Figure 11 (a) shows the analysis using video 1. For

the scaling factor 4, PSNR of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patchbased denoising is 33.644dB, 21.79dB, 19dB, 21.53dB, 32.21dB, and 23.98dB, respectively. Figure 11 (b) shows the analysis using video 2. For the scaling factor 4, PSNR of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 31.3646dB, 13.90dB, 20.66dB, 11.45dB, 21.52dB, and 24.09dB, respectively. The proposed method exhibited the greater value of the PSNR. Figure 11 (c) shows the analysis using video 3. For the scaling factor 4, PSNR of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 30.974dB, 12.619dB, 21.07dB, 10.3033dB, 20.4912dB, and 23.88dB, respectively. For all the scaling factors considered, the proposed method exhibited the greater value of the PSNR.

SDME: The analysis based on the SDME of the comparative methods is progressed using three videos and is depicted in figure 12. Figure 12 (a) shows the SDME analysis using video 1. For the scaling factor 4, SDME of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 45.277dB, 36.651dB, 38.0591dB, 43.0591dB, 44.7630dB, and 37.8764dB, respectively. The proposed method exhibited the greater value of the SDME than the existing methods. Figure 12 (b) presents the SDME analysis using video 2. For the scaling factor 4, the SDME of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 43.7399dB, 37.866dB, 39.294dB, 42.583dB, 43.216dB, and 41.239dB, respectively. Figure 12 (c) shows the analysis based on SDME using video 3. For the scaling factor 4, the SDME of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 53.7979dB, 29.75dB, 41.13dB, 41.8065dB, 41.965dB, and 43.866dB, respectively. The proposed method exhibited
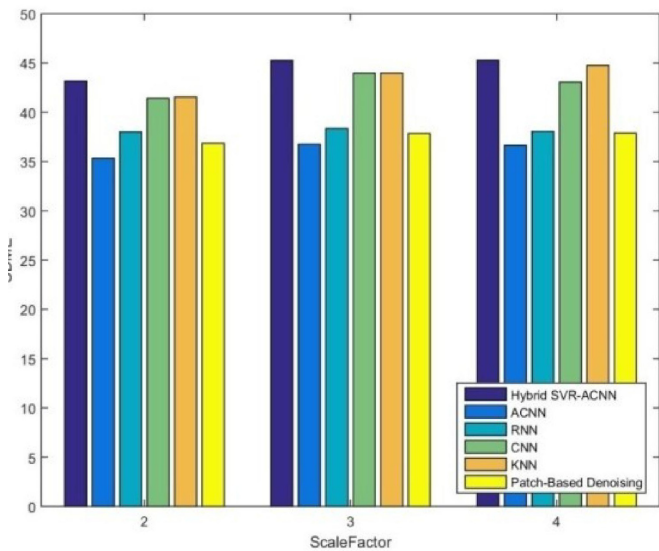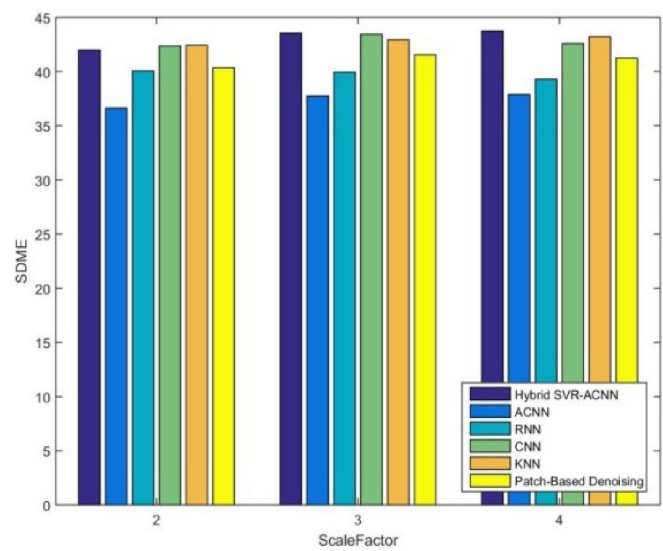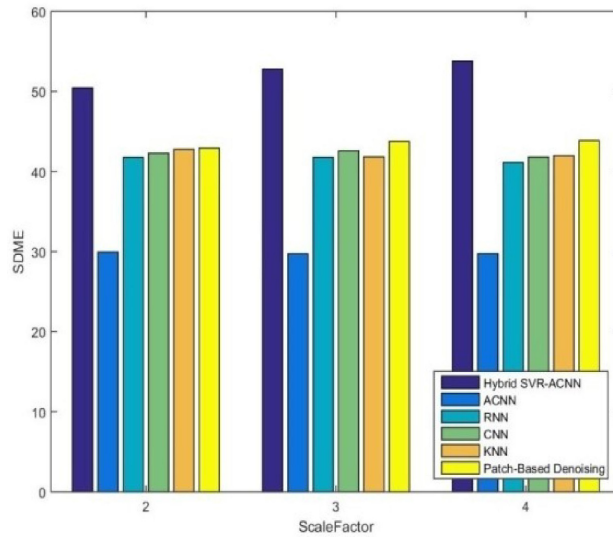
(a)
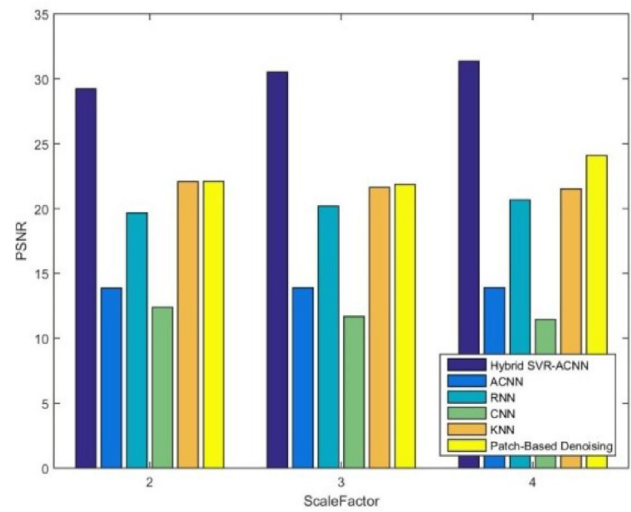
(b)
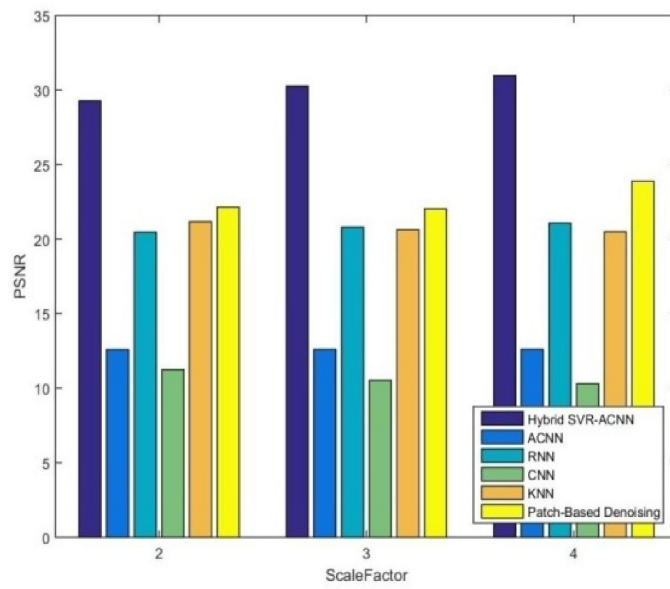
(c)

Figure 11. Comparative analysis based on PSNR (a) video 1 (b) video 2 (c) video 3
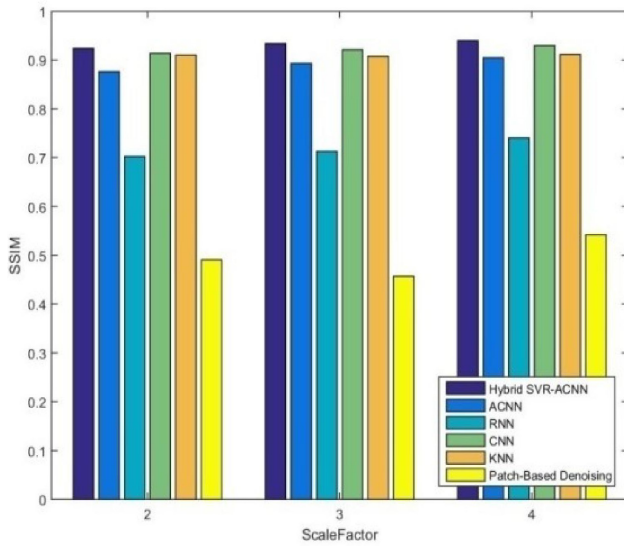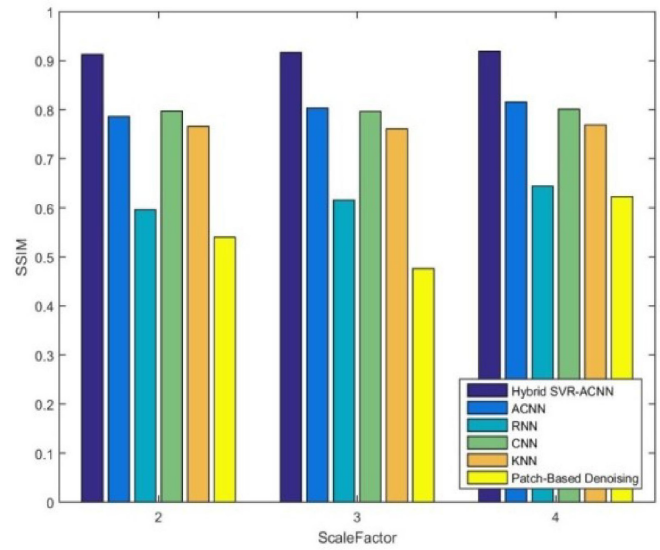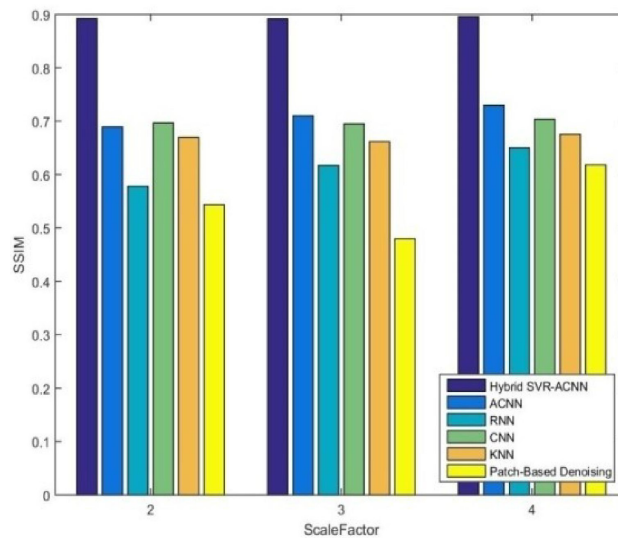

(a)

(b)

(c)

Figure 12. Comparative analysis based on SDME (a) video 1 (b) video 2 (c) video 3

(a)

(b)

(c)

Figure 13. Comparative analysis based on SSIM (a) video 1 (b) video 2 (c) video 3

Methods	PSNR (dB)	SSIM	SDME (dB)
Proposed Hybrid SVR-ACNN	33.64447	0.9398	45.2779
ACNN	21.79351	0.9045	36.6517
RNN	19	0.7404	38.0591
CNN	21.5312	0.9297	43.0591
KNN	32.21	0.9113	44.763
Patch-based denoising	23.98	0.5422	37.876

Table 1. Comparative analysis using the video super-resolution methods

the greater value of the SDME for all the videos irrespective of the scaling factor.

SSIM: The analysis of the comparative methods based on the SSIM is progressed using three videos, as depicted in figure 13. Figure 13 (a) shows the analysis using video 1. For the scaling factor 4, the SSIM of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 0.9398, 0.9045, 0.7404, 0.9297, 0.9113, and 0.5422, respectively. Figure 13 (b) shows the analysis using video 2. For the scaling factor 4, the SSIM of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 0.9190, 0.8157, 0.9439, 0.8007, 0.7688, and 0.6226, respectively. Figure 13 (c) shows the analysis using video 3. For the scaling factor 4, SSIM of the proposed hybrid SVRACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 0.8956, 0.7297, 0.6498, 0.7036, 0.6754, AND 0.6182, respectively. From figure 13, it is seen that the proposed method exhibited the greater value of the SSIM.

5.8 Comparative Discussion
The comparative analysis of the SR methods providing the maximum performance is given in table 1. The PSNR of the proposed hybrid SVR-ACNN, ACNN, RNN, CNN, KNN, and Patch-based denoising is 33.6444 dB, 21.79351 dB, 19 dB, 21.5312 dB, 32.21 dB, and 23.98 dB. The SSIM of ACNN and RNN is 0.9045, and 0.7404, whereas that of the proposed hybrid SVR-ACNN is 0.9398. The maximum SDME attained by the proposed hybrid SVRACNN is 45.2779dB when the existing KNN could provide the SDME of 44.763 dB. The proposed method acquired a greater value of PSNR, SSIM, and SDME, respectively.

6. Conclusion

Video Resolution image generated using the proposed model is a high-quality video that is applicable to the processes associated with high video applications. The hybrid model has been developed using the SVR model and the Actor critic neural network that is merged together based on the weighted average concept. The tuning of the ACNN is based on the newly devised algorithm, F-

based SCA, which possess the capacity to compute the weights based on the past records and holds a faster converging property that converges to the global optimal solution. The proposed model of video enhancement addresses the demerits of the existing SR methods, offering an effective SR methodology. The experimentation is performed using the Cambridge-driving Labeled Video Database (CamVid), and the effectiveness of the proposed method is analyzed based on PSNR, SDME, and SSIM, respectively. The proposed method outperforms the existing SR methods with the maximum PSNR of 33.6447dB, maximum SDME of 45.2779dB, and maximum SSIM of 0.9398, respectively.

References

[1] Kappeler, Armin., Yoo, Seunghwan., Dai, Qiqin., Aggelos, Katsaggelos, K. (2016). Video super-resolution with convolutional neural networks, *IEEE Transactions on Computational Imaging*, 2 (2) 109-122.

[2] Li, Kun., Zhu, Yanming., Yang, Jingyu., Jiang, Jianmin. (2016). Video super-resolution using an adaptive superpixel-guided auto-regressive model, *Pattern Recognition,* 51, p. 59- 71, 2016.

[3] Yang, Wenhan., Feng, Jiashi., Xie, Guosen., Liu, Jiaying., Guo, Zongming., Yan, Shuicheng. (2017). Video super-resolution based on spatial-temporal recurrent residual networks, *Computer Vision and Image Understanding*.

[4] Li, Dingyi., Wang, Zengfu. (2017). Video Super-Resolution via Motion Compensation and Deep Residual Learning, *IEEE Transactions on Computational Imaging*, 3 (4) 749-762.

[5] Amar, B., Deshmukh, Usha Rani, N. (2017). Fractional-Grey Wolf optimizer-based kernel weighted regression model for multi-view face video super resolution, *International Journal of Machine Learning and Cybernetics*, p.1-19, 23 December 2017.

[6] Li, Yawei., Li, Xiaofeng., Fu, Zhizhong. (2017). Modified non-local means for superresolution of hybrid videos, *Computer Vision and Image Understanding*, 2 December 2017.

[7] Huang, Yan., Wang, Wei., Wang, Liang. (2017). Video Super-Resolution via Bidirectional Recurrent Convolutional Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 99, p.1-1.

[8] Chen, Di., He, Xiaohai., Chen, Honggang., Wang, Zhengyong., Zhang, Yijun. (2016). Video superresolution using joint regularization, *In*: Proceedings of the IEEE 13th International Conference on Signal Processing (ICSP)p.668 - 672.

[9] Buades, Antoni., Lisani, Jose-Luis., Miladinovic, Marko. (2016). Patch-based video denoising with optical flow estimation, *IEEE Transactions on Image Processing*, 25 (6) 2573-2586.

[10] Wang, Jen-Wen., Chiu, Ching-Te. (2017). Video Super-resolution using Edge-based Optical Flow and Intensity Prediction, *Journal of Signal Processing Systems*, p.1-13.

[11] Zhang, Xinfeng., Xiong, Ruiqin., Ma, Siwei., Li, Ge., Gao, Wen. (2015). Video super-resolution with registration-reliability regulation and adaptive total variation, *Journal of Visual Communication and Image Representation*, 30, p.181-190, July 2015.

[12] Hiding Yunbo Rao, Chen, Leiting. (2012). A Survey of Video Enhancement Techniques, *Journal of Information and Multimedia Signal Processing*, 3 (1) 71-99, January 2012.

[13] Bhagya, H. K., Keshaveni, N. (2016). Review on video enhancement techniques, *International Journal of Engineering Science Invention Research & Development*, 3 (2) August 2016.

[14] Zhang, Z., Sze, V. (2016). Fast: Free adaptive super-resolution via transfer for compressed videos, arXiv preprint arXiv:1603.08968, 2016.

[15] Jin, Y., Bouganis, C. S. (2015). Robust multi-image based blind face hallucination, *In*: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, p. 5252–5260.

[16] Zhang, L., Zhang, H., Shen, H., Li, P. (2010). A super-resolution reconstruction algorithm for surveillance images, Signal Process., 90 (3) 848–859.

[17] Zhong, Y., Zhang, L. (2012). Remote sensing image subpixel mapping based on adaptive differential evolution, *IEEE Transactions on Systems, Man, and Cybernetics B*, Cybern, 42 (5) 1306–1329.

[18] Wallach, D., Lamare, F., Kontaxakis, G., Visvikis, D. (2012). Superresolution in respiratory synchronized positron emission tomography, *IEEE Transactions on Medical Imaging*, 31(2) 438–448.

[19] Zhang, J., Cao, Y., Zha, Z. J., Zheng, Z., Chen, C. W., Wang, Z. (2016). A unified scheme for uper-resolution and depth estimation from asymmetric stereoscopic video, *IEEE Transactions on Circuits and Systems for Video Technology*, 26 (3) 479–493.

[20] Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., Zhang, L. (2016). Image superresolution: The techniques, applications, and future, Signal Process, 128, p. 389–408.

[21] Freeman, W. T., Jones, T. R., Pasztor, E. C. (2002). Example-based superresolution, *IEEE Computer Graphics and Applications*, 22 (2) 56–65.

[22] Chang, H., Yeung, D.-Y., Xiong, Y. (2004). Super-resolution through neighbour embedding, *In:* Proceedings of the IEEE conference on computer vision and pattern Recognition, 1, 2004, p. 275– 282.

[23] Yang, J., Wright, J., Huang, T., Ma, Y. (2008). Image super-resolution as sparse representation of raw image patches, *In:* Proceedings of the IEEE conference on computer vision and pattern Recognition, 2008, p. 1–8.

[24] Timofte, R., De Smet, V., Van Gool, L. (2013). Anchored neighbourhood regression for fast example-based super-resolution, *In:* Proceedings International Conference on Computer Vision, 2013, p. 1920–1927.

[25] Dong, C., Loy, C. C., He, K., Tang, X. (2014). Learning a deep convolutional network for image super-resolution, *In: Proceedings European Conference on Computer Vision*, 2014, p. 184–199.

[26] Liu, C., Sun, D. (2014). On bayesian adaptive video super resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 346–360.

[27] Williams, R. J., Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. *Neural Computation*, 1 (2) 270–280.

[28] Mirjalili, Seyedali. (2016). SCA: A Sine Cosine Algorithm for solving optimization problems, Knowledge-Based Systems, 96, p. 120-133, March 2016.

[29] Motion-based Segmentation and Recognition Dataset taken from http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/, accessed on December 2017.

[30] Zhao, Dongbin., Wang, Bin., Liu, Derong. (2013). A supervised Actor–Critic approach for adaptive cruise control, *Soft Computing*, 17 (11) 2089–2099, November 2013.

[31] Karl, S., Ni, Truong Q., Nguyen. (2007). Image Superresolution Using Support Vector Regression, *IEEE Transactions on Image Processing,* 16 (6)1596 - 1610.

[32] Pawan, R., Bhaladhare, Devesh, C., Jinwala. (2014). A Clustering Approach for the -Diversity Model in Privacy Preserving Data Mining Using Fractional Calculus-Bacterial Foraging Optimization Algorithm, Advances in Computer Engineering, 2014, p.1-12.

[33] Daga, B. S., Ghatol, A. A. (2016). Detection of Objects and Activities in Videos using Spatial Relations and Ontology Based Approach in Video Database System, *International Journal of Advances in Engineering & Technology*, 9 (6) 640- 650.

[34] Diamantini, Claudia., Potena, Domenico. (2009). Bayes vector quantizer for class-imbalance problem, *IEEE*

---

[35] Schietroma, Mario., Piccione, Federica., Clementi, Marco., Emanuela Marina Cecilia, Sista, Federico., Pessia, Beatrice., Carlei, Francesco., Guadagni, Stefano., Amicucci, Gianfranco. (2017). Short-and long-term, 11–22 Years, results after laparoscopic nissen fundoplication in obese versus nonobese patients, *Journal of Obesity*.

[36] Di Furia, Marino., Della Penna, Andrea., Salvatorelli, Andrea., Marco, Clementi,. Stefano, Guadagni. (2017). A single thyroid nodule revealing early metastases from clear cell renal carcinoma: case report and review of literature, *International Journal of Surgery Case Reports*, 34, p. 96-99.

[37] Attaccalite, Claudio., Cannuccia, E, Grüning, M. (2017). Excitonic effects in third-harmonic generation: The case of carbon nanotubes and nanoribbons, *Physical Review B*, 95 (12).

[38] Palumbo, Paola., Miconi, Gianfranca., Cinque, Benedetta., Lombardi, Francesca., Cristina La Torre, Soheila Raysi Dehcordi, Galzio, Renato., Cimini, Annamaria., Giordano, Antonio., Maria Grazia Cifone, (2017). NOS2 expression in glioma cell lines and glioma primary cell cultures: correlation with neurosphere generation and SOX-2 expression, *Oncotarget,* 8 (15).

[39] Valsalan, Prajoona., Shibi O Manimegalai, Shine Augustine, P. (2017). Non invasive estimation of blood pressure using a linear regression model from the photoplethysmogram (PPG) signal, Perspectivas em Ciencia da Informacao, 22 (4).