Turkish Journal of Computer and Mathematics Education Vol.12 No.9 (2021), 3098-3108

Research Article

Efficient Video Super-Resolution Using Deep Convolutional Autoencoders

Rohita H. Jagdale[1], Dr. Sanjeevani K. Shah[2]

[1]Research Scholar (E &TC), Maharashtra Institute of Technology, Kothrud, Pune.
[2]Professor and Head, (PG-E & TC), Smt. Kashibai Navale College of Engineering, Vadgaon, Pune.

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: Video super-resolution is the most well-known area of research in computer science. A video super-resolution technique is commonly required to recreate high-resolution video from noisy, blurry, and low-resolution video. Super-resolution is used in many applications like biomedical image processing, computer vision, and satellite image processing. This paper proposes a deep convolution auto-encoder-based video super-resolution model, which is trained with high-resolution video frames. An autoencoder is an unsupervised neural network that learns how to minimize data through design and reconstructs the loss of as little data as possible. In the test model, low-resolution frames are extracted from the low-resolution video. These low-resolution frames are then passed to the proposed architecture, which is modeled using a convolution auto-encoder. Important features of frames are extracted using multiple convolutional layers and different filters in the encoder model. High-resolution frames are reconstructed using decoder by minimizing loss function using L1 regularization with backpropagation, and weight matrices are updated with Adam optimizer. The proposed model's efficiency is evaluated and contrasted with state-of-the-art PSNR, SSIM, BRISQUE, VIFP, and UQI techniques. The proposed autoencoder model shows excellent performance.

Keywords: Video Super-Resolution, Deep Learning, Autoencoder.

I. Introduction

Advanced imaging is taking a great part in daily life and continually requires better picture quality, a higher goal, and greater usefulness. High resolution (HR) implies however much data would be prudent inside a given size of the image. So, high-resolution recordings as a regulation offer significant or even basic data for different military and regular citizen applications like investigation screens, clinical imaging, and so on. Nonetheless, the utilization of better image sensors and optics is a costly and restricting method of intensifying pixel density inside the image. An elective methodology is to improve the spatial resolution of low-resolution images. This can be accomplished by expanding the size of the sensor chip. Expanded chip size results in high capacitance, which thus hampers the general charge move rate. Likewise, the expanded chip size will cause reduced bundling of imaging gadgets, which will likewise add expense to that gadget. Subsequently, this methodology is not a powerful arrangement. Signal processing explicitly picture preparing is utilized to conquer these restrictions of the sensors and optics fabricating innovation. A modest and powerful answer to get a high resolution (HR) image from low resolution (LR) pictures is picture preparing techniques. This sort of image upgrading is called picture super-resolution (SR). The image handling approach offers the advantage of the usage of the current accessible low-resolution imaging framework. Because of weaknesses like camera cost, power, memory size, and restricted data transfer capacity, it isn't generally conceivable to get high-resolution pictures.

The super-resolution (SR) is the reproduction of outwardly satisfying high-resolution (HR) images from low-resolution (LR) images. For quite a long time, SR has pulled in broad consideration, and it expects to recover excellent images from low-cost imaging gadgets and some restricted ecological assets. Hence, growing better techniques for Super-Resolution is fundamental, explicitly thinking that HR pictures can be given as proof in a courtroom, as a piece of clinical records, and for the satellite caught images for acknowledgment. In this way, getting a high-resolution picture is a fundamental thought process of the super-resolution. Super Resolution calculations can be comprehensively characterized into two classes depending on the Frequency domain and Special domain [13]. Frequency domain methods are classified into two classes wavelet transform, and Fourier transform. Special domain algorithms are classified into three classes "Interpolation based, Reconstruction based and Learning based methods"[16]-[19].

II. Related work

The literature-based on several images used to perform super-resolution is described in this section. Various methods for super-resolution are depended on several low-resolution input images. The methods are classified according to single image [1]-[2], multiple images [3]-[5], and video super-resolution [6]-[10].

A. Single Image Super-Resolution

The "single-image super-resolution (SISR) methods" based on one low resolution (LR) image are also known as a one-input-one-output system. SISR algorithms are applied only on a single image and generate a single high resolution (HR) image as output. SISR can be achieved through "Interpolation based methods, Reconstruction based methods, and Learning based methods" [11]-[12].

Jaehwan Jeon et al. [1] proposed "Single Image Super-Resolution by Modifying Sampling Positions." Using variable sampling positions, a novel single-image super-resolution (SR) approach is introduced. This process calculates a sampling position correction vector (SPCV) from the frequently sampled data centered on the image's local gradient. The non-uniformly sampled data obtained by the SPCV is upgraded using the steering kernel regression method to retain the edge and remove strange artifacts in the SR process. Thanks to the coherent nature of vector sampling positions and directionally adaptive kernel regression, the method does not need to solve any partial differential equations (PDEs) or use an iterative technique. It can quickly be extended to multiple pre and post-processing methods for further enhancement.

Single Image Super-Resolution Based on Feature Enhancement described Shiyao Suo et al. [2]. In this, an effective single image super-resolution (SISR) method is proposed, which restores a high-resolution (HR) image through the feature enhancement mapping matrices and detail enhancement mapping matrices. The strategy is based on changed moored neighborhood relapse to get more exact features and better mapping matrices. In the system, the linear mapping matrices are twice organized. In the principal stage, interjected LR image features and HR image features are removed to learn include improvement mapping matrices. At that point, these learned grids are utilized to upgrade the extricated features in the subsequent stage, taken into detail improvement mapping matrices from upgraded LR picture features and HR picture patches. With the two-stage learning methodology, more exact mapping matrices can be acquired, and along these lines, better SR results can be accomplished.

B. Multiple Images Super-Resolution

Multiple images super-resolution (MISR) takes multiple low-resolution (LR) image versions and try to merge them to retrieve the high-resolution (HR) image version data. Various image super-resolution reproduces high-resolution pictures by consolidating a few low-resolution images. The MISR [3]-[5] mainly uses reconstruction based methods.

Zehua Lyu, et al.[3] proposed restoration of multi-image super-resolution utilizing a novel model of degradation. A significant topic in image processing is multi-frame Super-Resolution Reconstruction (SRR). The SRR model may be split into two sections using the Bayesian framework: the model of degradation and the prior model. Many scholars are presently working on the prior models, and several impressive previous models have been suggested. A perfect and simplistic model, though, is the widely used deterioration model. In the deterioration process, it only takes into consideration the noise distortion. In the degradation process, though, there is a lot of details lost. Therefore, a new degradation model is proposed in this, which aims to recreate a better picture of high resolution and missing data.

Multiple-image super-resolution defines multiple image super-resolution utilizing both reconstruction optimization and deep neural network[4]. In this, a revolutionary multi-image super-resolution approach (MISR) was proposed, cascading a reconstruction-dependent SR and a filter for the elimination of objects based on a three-layer deep neural network (DNN). In specific, several images with sub-pixel offsets are taken by the reproduction-based technique as input, and one high-resolution picture is output. A three-layer coevolutionary neural organization is introduced to remove the residues induced by the incorrectly presented problem of remaking and sharpening the edges. Using the L1-norm regularization term in the proposed algorithm to limit the reconstruction process for fast convergence, the conjugate-gradient algorithm is used. Ringing objects are unavoidable with these restored high-definition photographs because of the ill-position of the super-resolution. They devise a neural network of three-layer convolutional layers to eliminate the ringing artifacts and recover the rough edges. The device obtains both from the different low-resolution inputs and the data derived from the three-layer neural network, achieving superior performance that exceeds the state-of-the-art SR approaches.

Seokhwa Jeong, et al.[5] suggested utilizing locally directional self-similarity multi-frame example-based super-resolution. Multi-frame super definition reconstructs a high-resolution image from several low-resolution video clips. The algorithm consists of three steps: (i) the local search region definition using motion vectors for the optimal patch, (ii) the optimal patch adaptive collection based on the lower solution image degradation model, and (iii) the combination of the optimal patch and the reconstructed image. Based on the lower solution image degradation model, the usage of directionally adaptive patch selection, the algorithm will eliminate interpolation objects. Also, it is possible to create super-solved images without distortion between consecutive pictures. The strategy provides a significantly enhanced implementation of super-resolution over existing methods.

C. Video Super Resolution

Unified single image and video super-resolution via de-noising algorithms [6] describe a clear and versatile super-resolution architecture that can be applied to single images and easily expanded to video[14]. This relies on the perception that a range of techniques is overseen and correctly handled by de-noising photographs and recordings. It

takes advantage of the Plug-and-Play-Prior paradigm and the de-noising (RED) regularization method that expands it and shows how to use such de-noisers to use a single formulation and system to cope with the SISR and VSR issues. This way, it profits from current de-noising image/video algorithms' reliability and efficacy while solving even more complex problems[13]. More precisely, the VBM3D video denoiser binding obtains a highly competitive VSR algorithm free of motion estimation, exhibiting a propensity to high-quality performance and quick operation.
Xiaoting Du et al. [7] suggested a dense-connected residual network for video super-resolution. A dense-connected residual network (DCRnet) is a modern approach for super-resolution video to overcome the above disadvantages. By taking advantage of the hierarchical properties of all coevolutionary layers, "the DCR net may retain the low-frequency quality of motion-compensated frames and promote the restoration of high-frequency data. A dense-connected residual block (DCRB) was explicitly suggested as a critical factor". The performance of one DCRB is the compact concatenation of all previous DCRB features and each of the current DCRB's residual block features[11].
Zhi Liu, et al.[8] define a modern low bit-rate coding scheme for an ultra-high quality video focused on super-resolution reconstruction. A modern method of low bit-rate coding is recommended. On the encoder hand, "the video sequences are down-sampled to reduce the bit-rate of the encoded video stream remarkably." Meanwhile, to recreate high-definition film, an enhanced super-resolution reconstruction algorithm is added to the encoding hand's decoded data. The accuracy of the restored video is ensured thanks to the use of the super-resolution restoration algorithm. BD-PSNR experimental findings indicate that under the low bit rate, the scheme is stronger than HEVC.
HDR Video Super-Resolution was suggested by Seiya Umeda et al. [9] for potential video coding. Modern coding techniques in Future Video Coding are evolving (FVC). To show the state of the art coding results in ultra-high definition video coding, a novel video coding method that incorporates FVCera coding resources with super-resolution is defined. Experimental findings indicate that in both SDR and HDR photos, the proposed method outperforms the convention system.
Wenjing Yu et al. [10] explained the "Super-Resolution Reconstruction Video Image based on the Enhanced Glowworm Swarm Optimization Algorithm. An effective video image super-resolution[12] reconstruction model is created for the pixel resemblance of continuous multi-frame image sequences in video super-resolution reconstruction, which is transformed into an optimization problem low-resolution image pixel sequence to a high-resolution image pixel sequence. Since the basic Glowworm Swarm Optimization (GSO) algorithm is easy to split through the severe value oscillation and the local optimum, the global optimal individual impact factor and the local optimal individual impact factor are applied to the location update technique, the volatilization and gain coefficient of firefly fluorescence are improved. The super-resolution reconstruction characteristics redefine an Improved Glowworm Swarm Optimization (IGSO) The algorithm's swarm data, firefly's luciferase, and position update equation; the optimization objective function criterion is set".

III. Deep learning for Video Super-resolution

In this section, the general structure of auto-encoder, proposed architecture of autoencoder based video super-resolution model, loss function, backpropagation, regularization, and optimization function is described.

A. Overview of Autoencoder

An autoencoder is an artificial neural network that learns in an unsupervised manner [20]. It consists of three components input layer (encoder), code, and output layer (decoder), as shown in Fig. 1. The encoder's function is to compact the data to a lower dimension in such a way that only the most influential functions, such as Principle Component Analysis (PCA), remain, and the decoder's role is to rebuild the data as near as it can get to the original from the lower-dimensional representation. Input layer maps input x with h as

$$h = \sigma(Wx + b)$$

Where "h is code, σ is an element-wise activation function, x is input, W is a weight matrix, and b is a bias vector. Weights and biases are adjusted arbitrarily and then updated iteratively during training through backpropagation".

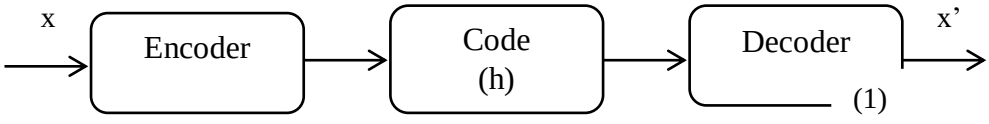


Figure 1.Basic model of the autoencoder

The output stage maps "h to the reconstruction x."

$$x' = \sigma'(W'h + b')$$

Where "h is code, $\sigma'$is an element-wise activation function, x' is reconstructed output, W' is a weight matrix, and b' is a bias vector."
The main objective function of an autoencoder is to minimize loss.

$$L(x, x') = \left\| x - x' \right\|^2 = \left\| x - \sigma'(W'(\sigma(Wx + b)) + b') \right\|^2$$

B. Proposed Model of Video Super-resolution

Figure 2 shows the proposed model of video super-resolution. It consists of input low-resolution frames, auto-encoder-based super-resolution models, and reconstructed high-resolution output frames. In the image restoration task, they are normally implemented to reduce reconstruction errors by selecting the most suitable filters. To extract attributes, once they are qualified in this role, they can be added to any input. Convolutional autoencoders are extractors of general-purpose functionality that disregard the 2D image form, unlike general autoencoders. The frame must be unrolled into a single vector in autoencoders, and the network must be constructed following the restriction of the number of inputs.

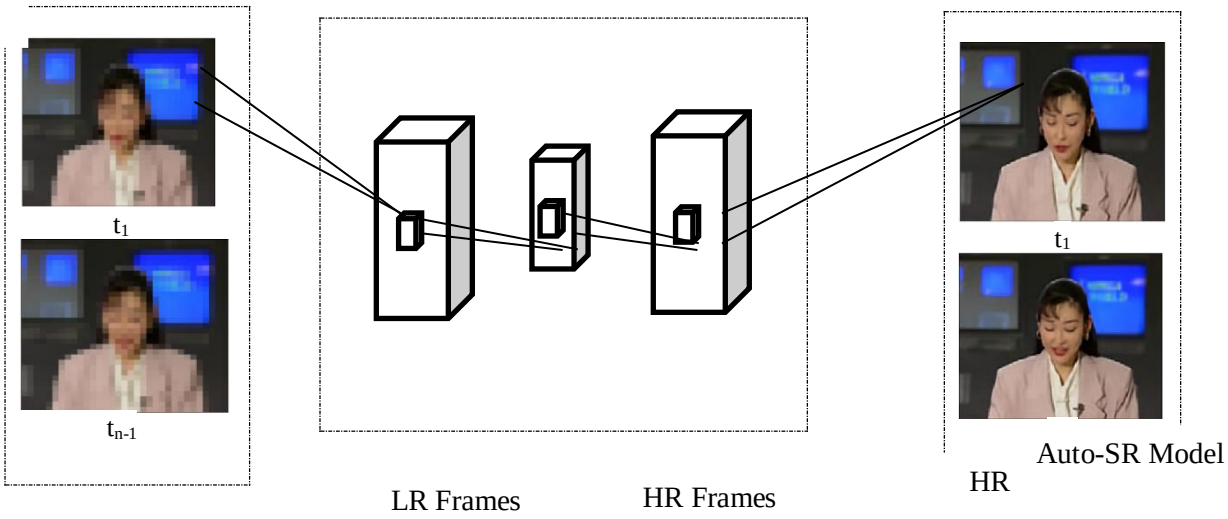


Figure 2.Proposed flow diagram of Video SR Model

UCSD and Xiphvideo datasets [23] are used for training and testing. The proposed architecture is shown in Fig. 3, with five convolutional layers in the encoder and five convolutional layers in the decoder. Encoder extracts the important features of frames and removes redundant information. The size of the feature map is smaller than the input frame size. The residual connection is used from encoder to decoder to diminish the final reconstructed frame's lossy results. The number of filters is increased in two folds. To adjust the weight matrices, backpropagation is used, and weights are optimized with adam optimizer.

Input low resolution (LR) I^{LR} and output high-resolution frame I^{HR} both have C color channels, due to which low-resolution frames are represented as H x W x C. The upscaling factor is considered r, so the high-resolution frame has rH x rW x C. To recover the high resolution frames I^{HR}, five layers convolutional network is used.

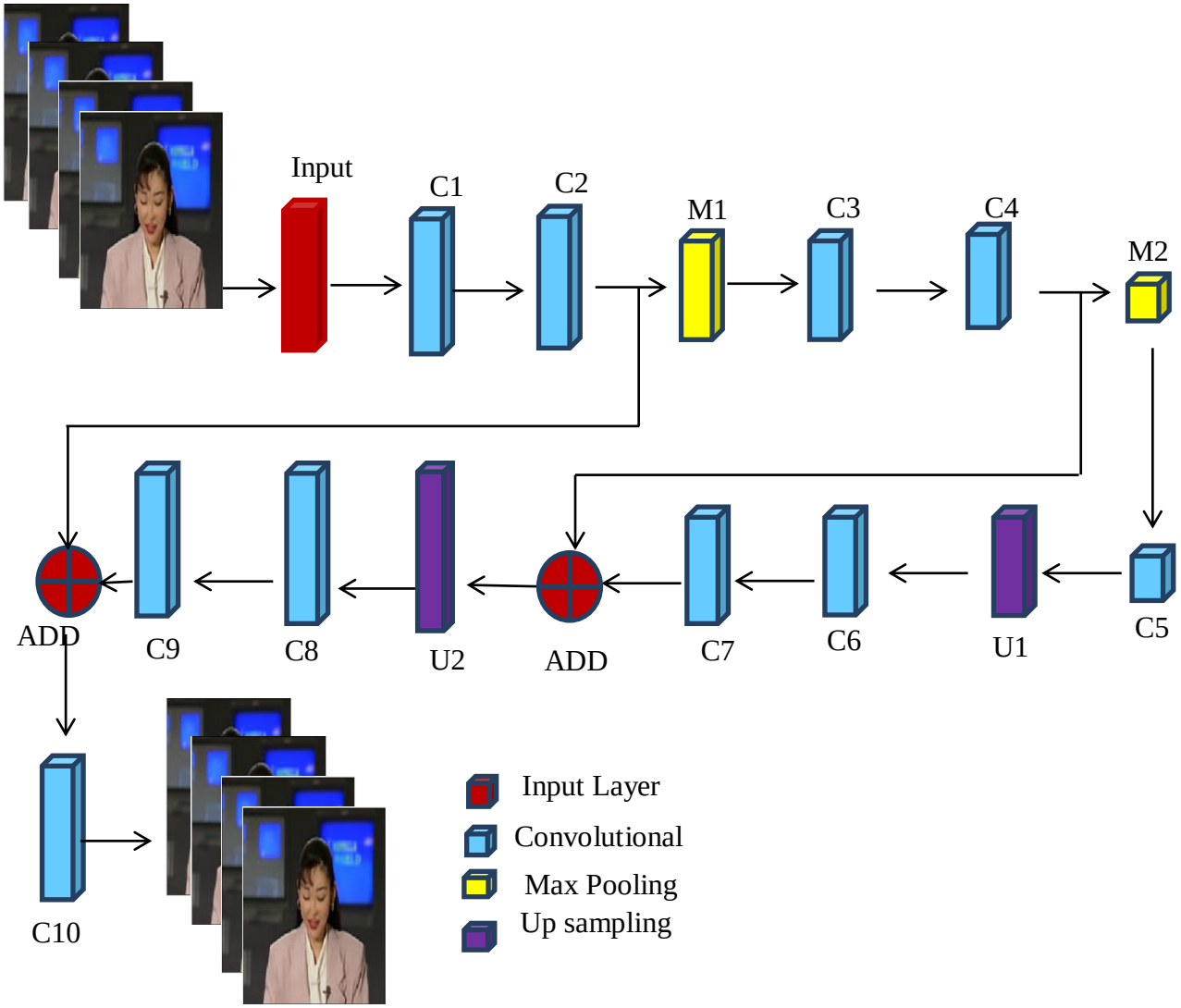


Figure 3.Model of proposed auto-encoder for feature extraction

In the above network of L layers, the first L-1 layers are

$$f^{l}(I^{LR};W_1,b_1)=\phi(W_1 * I^{LR}+b_1) \quad (4)$$

$$f^{l}(I^{LR};W_{1:l},b_{1:l})=\phi(W_l * f^{l-1}(I^{LR})+b_l) \quad (5)$$

Where, W_l, b_l are learnable weights and biases

Convolutional Layer: A convolution layer determines a window through which we inspect a part of the image and subsequently scans the whole image looking through this window. This window is often called a filter since it creates an output picture that focuses exclusively on the image regions that showed the element for which it was looking. The performance of a convolution is referred to as a map of features.

Max pooling Layer: After the convolutional layer, the pooling layer is normally mounted. The pooling layer's usefulness is to decrease the input volume's spatial dimension for the next layers. "The pooling layer takes an input volume of size $W_l \times H_l \times D_l$. The two hyperparameters used are Spatial Extent F and Stride length S. The output volume is $W_{l+1} \times H_{l+1} \times D_{l+1}$ where $W_{l+1}=(W_l-F)/S+1$, $H_{l+1}=(H_l-F)/S+1$ and $D_{l+1}=D_l$. The pooling layer computes a fixed function, and in our case the max function, so there are no learnable parameters".

Up-sampling Layer: Up-sample layer is a simple layer with no weights that doubles the input dimensions. Different interpolation techniques are used in up-sampling to obtain a coarse high resolution (HR) frame. There are different ways to use up-sampling in a model, like pre-up-sampling, post upsampling, progressive upsampling, and iterative up and downsampling. In the proposed model, pre-sampling is used, in which a low-resolution frame is first upsampled with bilinear interpolation followed by a convolution layer.

C. Training objective Function

Mean square error is used as regression loss function as it is fastest. It is measured as the average squared difference between reconstructed and actual observations. In training, minimize loss function between the original frame and reconstructed frame as,

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (6)$$

Where y_i is the original frame and yi^ is the reconstructed frame

D. Regularize cost function

L1 and L2 are the two types of regularizers mostly used. These update the general expense feature as a regularization term is named by introducing another penalty term. The importance of weight matrices decreases due to the regularization concept since it implies that a neural network with smaller weight matrices contributes to simplified models. It would also help to minimize over-fitting by minimizing. In our model, L1 regularization is used; it is useful when trying to compress our model. We penalize the absolute value of the weights.λ is regularization parameter and its set as 0.001. It is the hyperparameter whose value is optimized for better results.

$$Cost\_Function = MSE + \lambda\sum\|W\| \quad (7)$$

E. Adaptive moment estimation (Adam) optimizer

A mixture of Adagard and RMSProp is an adaptive moment estimation (Adam) that works well in online and non-stationary environments. As in Adagard, it applies the exponential moving average of gradients to scale the learning rate instead of a standard average. It holds the average of previous gradients exponentially decaying. It is mathematically effective and needs relatively little memory. "Adam first updated the exponential moving averages of the gradient *(m_t)* and squared gradient *(v_t)*, which is an estimation of the first and second moment.

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$$
$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (9)$$

Where, m_t and v_t are an estimate of the first and second moment, Hyperparameters $\beta 1$is $\epsilon[0,1)$the expo[illegible]y rate for the first moment estimates, $\beta 2$is $\epsilon[0,1)$the exponential decay rate for the second-moment estim[illegible]f m_t and v_t.

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \quad (8)$$

Update the parameters

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (11)$$

$$\theta_{t+1} = \theta_t - \frac{\eta\, m_t}{\sqrt{v_t}+\varepsilon} \quad (12)$$

Where η is learning parameter, θ_t is the weight parameter, and ε is a very small number to prevent any di[illegible] zero in the implementation. Learning rate was kept at a fixed value of 0.001, $\beta_1=0.9$, $\beta_2=0.999$,$\varepsilon=10^{-8}$. The model is trained for up to 45 epochs.

IV. Results and Discussion

Experimentation conducted for proposed architecture, various design metrics used for subjective and objective analysis, and a dataset used for training and testing the proposed model are explained in this section. Performance assessment of the proposed model is carried out by comparing various metrics with other state-of-art techniques.

A. Performance Evaluation Metrics

Performance parameters used by "Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Visual Information Fidelity in Pixel Domain (VIFP) and Universal Quality Index (UQI) to measure Universal quality index (UQI),"

1.Peak signal-to-noise ratio (PSNR): It is an engineering concept for the ratio between a signal's highest potential strength and the noise corrupting power that influences its representation's fidelity. PSNR is generally represented in terms of the logarithmic decibel scale since certain signals have a very broad dynamic spectrum.

PSNR most conveniently describes the mean square error. "Given a noise-free m×n monochrome image I and its noisy approximation K, MSE is defined as:

$$\text{MSE}=\frac{1}{mn}\sum_{i}^{m-1}\sum_{j}^{n-1}[I(i,j) - K(i,j)]^2 \quad (13)$$

The PSNR (in dB) is defined as:

$$\text{PSNR} = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (14)$$

Here, MAX_I is the maximum possible pixel value of the image. When the pixels are represented using 8 bi[illegible] pe[illegible] sample, this is 255".

2. **Structural Similarity Index Measure (SSIM)**

A novel tool for calculating the resemblance between two pictures is the Structural Similarity (SSIM) index. As long as the other picture is assumed to be of perfect quality, the SSIM index can be interpreted as a quality indicator of

one of the photos being compared. To calculate the deterioration of a picture structure, SSIM uses a changed calculation of the spatial similarity between the reference pixels and the test image.

$$\text{SSIM (x, y)} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (15)$$

Where "μ_x the average of x; μ_y the average of y; σ_x^2 the variance of x; σ_y^2 the variance of y; σ_{xy} the covariance of x and y; $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$ two variables to stabilize the division with weak denominator; L the dynamic range of the pixel-values; K_1= 0.01 and K_2=0.03 by default;

3. Visual information fidelity in pixel domain (VIFP): The visual information fidelity (VIF) index is an objective quality metric that gives very accurate image similarity scores but at the cost of very high computational complexity".

The reference image information is

$$I(\vec{C}^N; \vec{E}^N \mid s^N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{M}\log_2\left(1+\frac{s^2{}_i\lambda_k}{\sigma_n^{\,2}}\right)$$

The test image information is

$$I(\vec{C}^N; \vec{F}^N \mid s^N) = \frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{M}\log_2\left(1+\frac{g_i^{\,2}s^2{}_i\lambda_k}{\sigma_v^{\,2}+\sigma_n^{\,2}}\right)$$

The visual fidelity measure is

$$\text{VIF} = \frac{\sum_{j\in subbands} I(\vec{C}^{N,j}; \vec{F}^{N,j} \mid s^{N,j})}{\sum_{j\in subbands} I(\vec{C}^{N,j}; \vec{E}^{N,j} \mid s^{N,j})}$$

B. Results

The performance of the proposed model is tested and evaluated on standard datasets xiph and UCSD. Fig. 4 and Fig. 5 show the model's performance for Epoch vs. Loss and Epoch vs. accuracy.

Total epochs are 100, and early stooping conditions appeared at epoch 67 with minimum loss0.0022 and maximum accuracy of 90%.

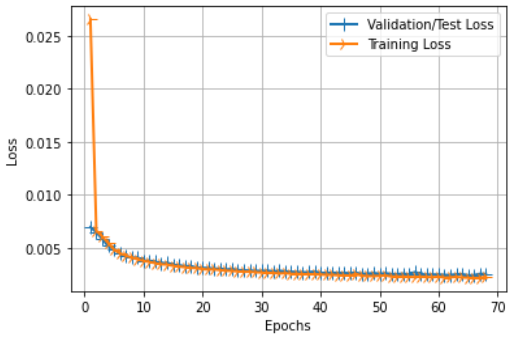


Figure 4. Simulation results of Epochs vs. Loss

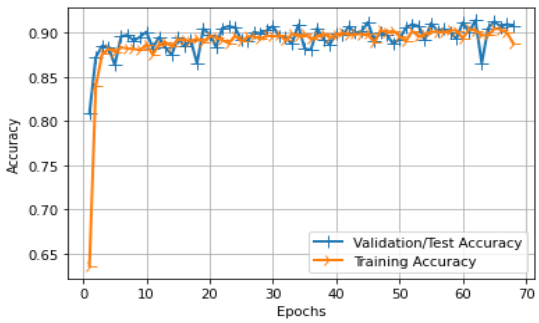
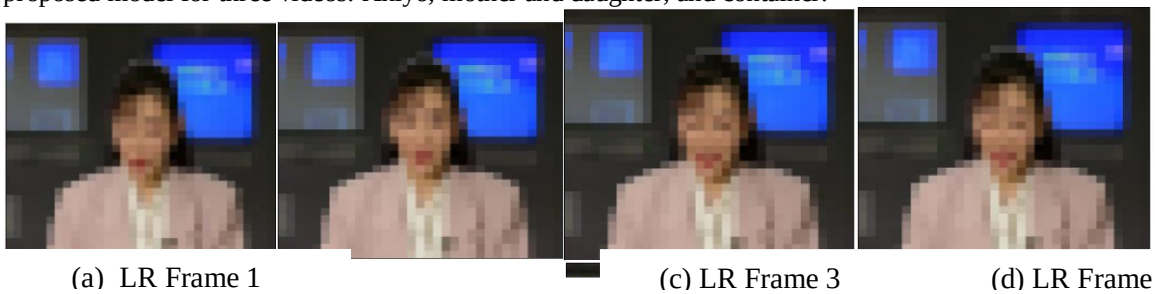


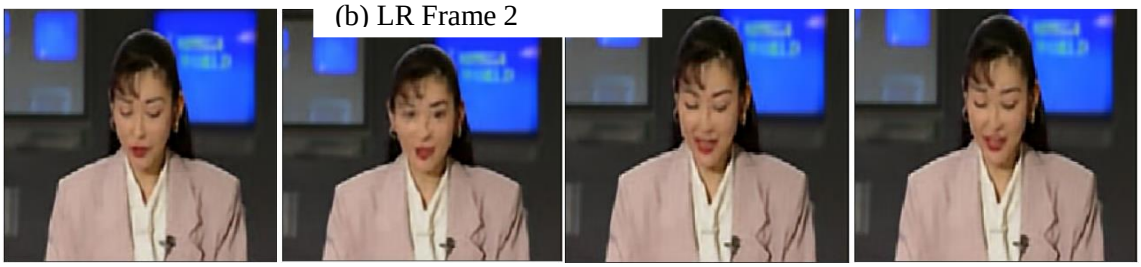
Figure 5. Simulation results of Epochs vs. Accuracy

C. Experimental results

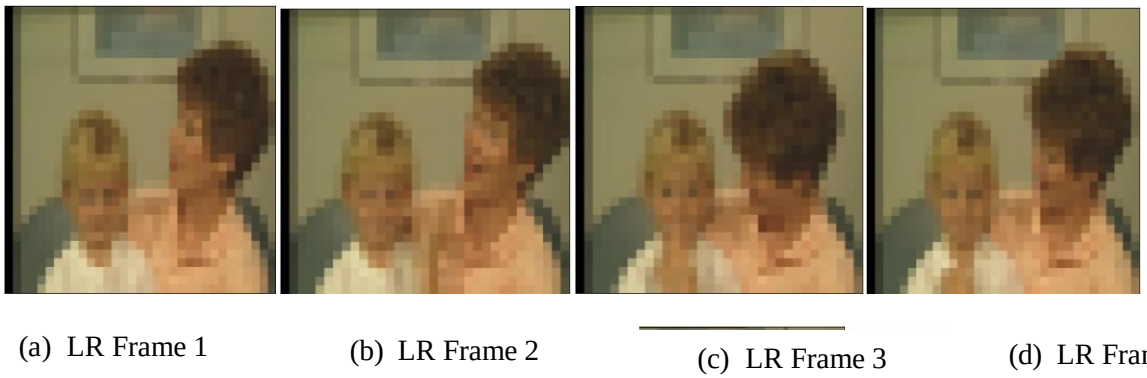
Figures 6, 7, and 8 show the visual quality of input lov
proposed model for three videos: Akiyo, mother and daughter, and container.



(a) LR Frame 1 (b) LR Frame 2 (c) LR Frame 3 (d) LR Frame 4

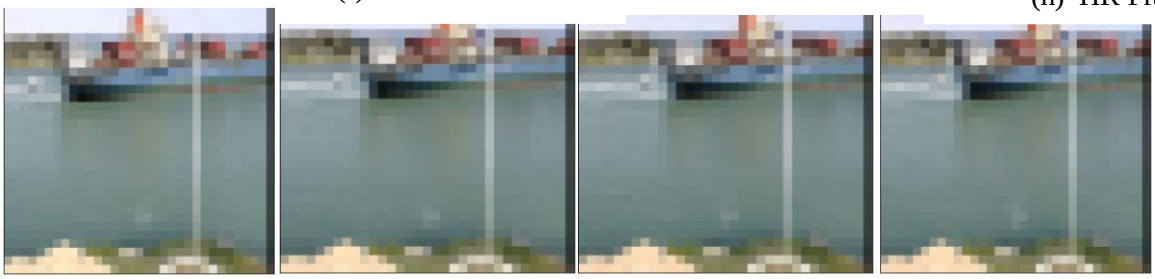


(e) HR Frame 1 (f) HR Frame 2 (g) HR Frame 3 (h) HR Frame 4

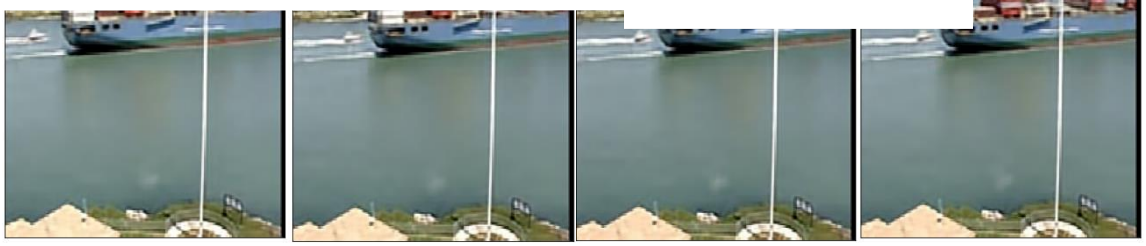


(a) LR Frame 1 (b) LR Frame 2 (c) LR Frame 3 (d) LR Frame 4

(e) HR Frame 1 (f) HR Frame 2 (g) HR Frame 3 (h) HR Frame 4



(a) LR Frame 1 (b) LR Frame 2 (c) LR Frame 3 (d) LR Frame 4



(e) HR Frame 1 (f) HR Frame 2 (g) HR Frame 3 (h) HR Frame 4

ntainer video proposed A

The proposed model's performance is compared with existing super-resolution algorithms with different parameters like PSNR, SSIM, VIFP, and UQI. PSNR of the proposed algorithm is improved by 6.8dB, 6.95dB,7.27dB, 7.03dB, and 1.42dB than Vandellwalle, Marcel, Lucchese, Keren, and Gholamreza algorithms, respectively. SSIM of the proposed algorithm is improved by 0.0562, 0.056, 0.058, 0.0558 and 0.0249 than Vandellwalle, Marcel, Lucchese, Keren and Gholamreza algorithms. VIFP of the proposed algorithm is improved by 0.2979, 0.2967, 0.2993, 0.2956, and 0.1209 than Vandellwalle, Marcel, Lucchese, Keren, and Gholamreza algorithms, respectively. UQI of the proposed algorithm is improved by 0.1499, 0.1479, 0.1475, 0.1475, and 0.068 than Vandellwalle, Marcel, Lucchese, Keren, and Gholamreza algorithms respectively.

Table 1: "Overall Statistical Performance of Proposed Auto-SR over other conventional models in terms of PSNR, SSIM, VIFP, and UQI for Akiyo.mp4, mother & daughter.mp4 and container.mp4."

Performance Measures	**Vandellwalle [19]**	**Marcel [19]**	**Lucchese [18]**	**Keren [17]**	**Gholamreza [21]**	**Proposed Auto-SR**
Akiyo.mp4						
PSNR	30.70	30.55	30.23	30.47	36.08	36.80
SSIM	0.9435	0.9437	0.9417	0.9439	0.9748	0.9997
VIFP	0.5596	0.5608	0.5582	0.5619	0.7366	0.8575
UQI	0.8205	0.8225	0.8164	0.8229	0.9024	0.9704
Mother and Daughter.mp4						
PSNR	28.98	28.97	28.95	28.98	34.30	36.88
SSIM	0.9097	0.9065	0.9016	0.9096	0.9450	0.9998
VIFP	0.4681	0.4686	0.4276	0.4731	0.6110	0.555
UQI	0.7361	0.7224	0.7122	0.7364	0.8568	0.9791
Container.mp4						
PSNR	24.48	24.37	24.18	24.44	28.71	34.02
SSIM	0.8339	0.8355	0.8299	0.8372	0.9134	0.9996
VIFP	0.3081	0.3113	0.3101	0.3149	0.4736	0.9421
UQI	0.6351	0.6392	0.628	0.6431	0.7848	0.9614

Figure 9 shows performance analysis with PSNR and SSIM of three different videos with other methods. It shows that the proposed model shows better PSNR and SSIM in all three videos. The first and second videos, akio.mp4 and mother and daughter.mp4, consist of static background with local motion. The third video, container.mp4, consists of dynamic background with global motions.

Figure 10 shows performance analysis with visual image fidelity in pixel domain (VIFP) and Universal quality index (UQI) of akio.mp4, mother, and daughter.mp4, and container.mp4. Our model shows significant improvement as compared to other methods.

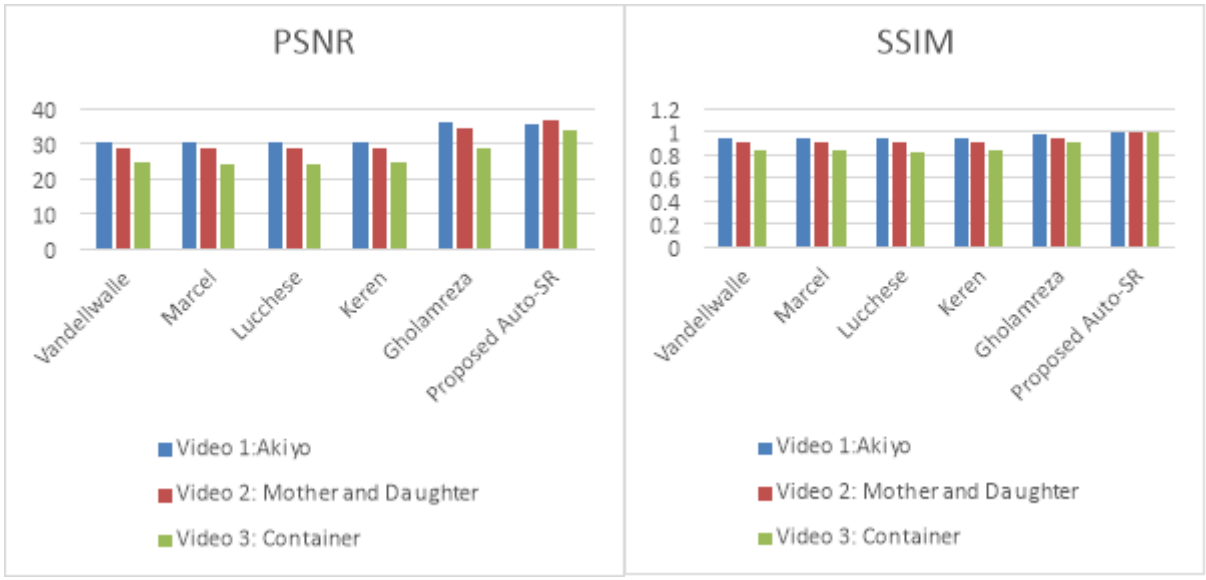


Figure 9.Graphical analysis of performance parameters PSNR (Left Side) and SSIM (Right Side)

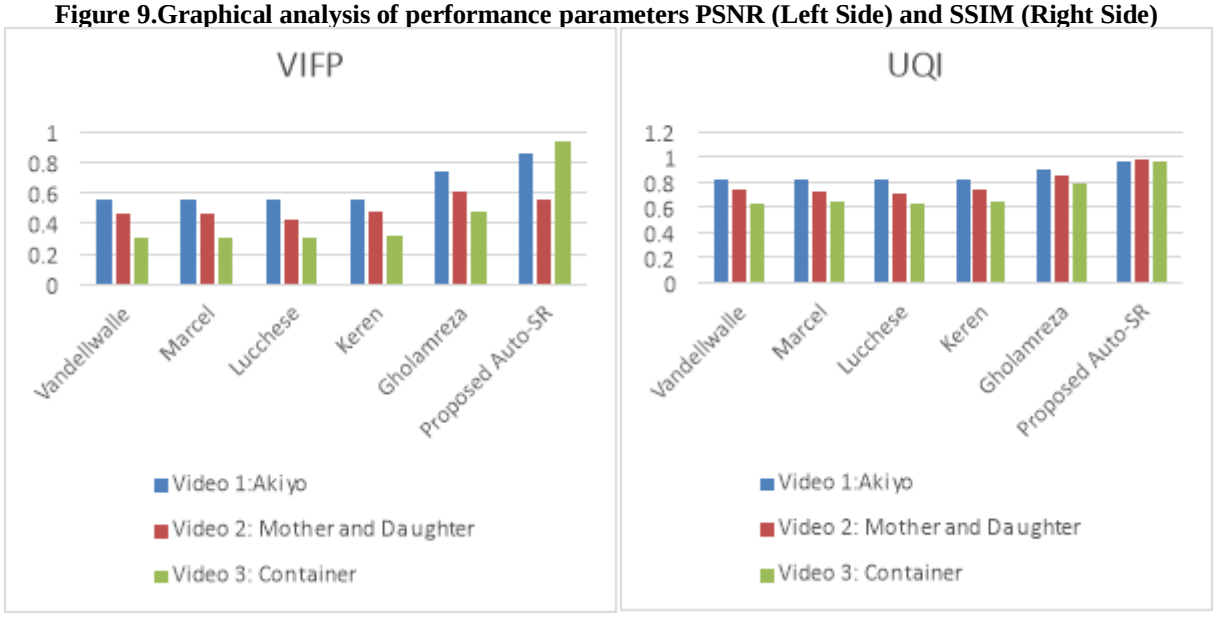


Figure 10 "Graphical analysis of performance parameters VIFP (Left Side) and UQI" (Right Side)

Conclusions

This paper proposes an auto-encoder-based super-resolution video model. Super-resolution video is a critical field of research for restoring high-resolution videos from noisy, distorted, and low-resolution videos in different applications. This paper describes a deep learning approach for video resolution enhancement, in which HD videos from the xiph dataset are used for training the model. The model gives 90% training accuracy and 91% testing accuracy with an MSE loss of 0.0022. Autoencoders are used to reconstruct input; it consists of an encoder for feature extraction and a decoder for reconstruction. Performance of video frames is evaluated with performance evaluation parameters PSNR, SSIM, VIFP, and UQI and compared with the state of the art methods. The proposed autoencoder based method gives a significant improvement in all four parameters.

References

1. Jaehwan Jeon, "Single Image Super-Resolution by Modifying Sampling Positions," IEEE International Conference on Consumer Electronics, pp. 258-259, 2014.
2. Shiyao Suo, "Single Image Super Resolution Based on Feature Enhancement," 2nd International Conference on Image, Vision and Computing, pp.473-477, 2017.
3. Zehua Lyu, "Multi Image Super Resolution Reconstruction Using A Novel Degradation Model," IEEE Transactions on Acoustics, Speech, and Signal Process, pp. 287-291, 2014.
4. Jie Wu, "Multiple-Image Super Resolution Using Both Reconstruction Optimization And Deep Neural Network," IEEE Transactions on Acoustics, Speech, and Signal Process, pp. 1175-1179, 2017.

5. Seokhwa Jeong, “Multi-Frame Example-Based Super-Resolution Using Locally Directional Self-Similarity,” IEEE Transactions on Consumer Electronics, Vol. 61, No. 3, August 2015.
6. Alon Brifman, “Unified Single-Image and Video Super-Resolutionvia Denoising Algorithms,” IEEE Transactions on Image Processing, 2019.
7. Xiaoting Du, “Dense-Connected Residual Network for Video Super-Resolution,” IEEE International Conference on Multimedia and Expo, 2019.
8. Zhi Liu, “A New Low Bit-Rate Coding Scheme for Ultra High Definition Video Based on Super-Resolution Reconstruction,” IEEE International Conference on Computer and Communication Engineering Technology, 2018.
9. Seiya Umeda, “HDR Video Super-Resolutionfor Future Video Coding,” IEEE International Conference on Computer and Communication Engineering Technology, 2018.
10. Wenjing Yu, “Super Resolution Reconstruction of Video Images Based on Improved Glowworm Swarm Optimization Algorithm,” 3rd IEEE International Conference on Image, Vision and Computing, 2018.
11. Deyun Wei, “Image super-resolution reconstruction usingthe high-order derivative interpolation associated with fractional filter functions,” IET Signal Processing, 2016.
12. Fraedreric Champagnat, “A Fourier Interpretation of Super-Resolution Techniques,” IET Signal Processing, 2016.
13. Said Assous, “High resolution time delay estimation using sliding discrete Fourier transform,” Digital Signal Processing, 2012.
14. G. Anbarjafari, S. Izadpanahi and H. Demirel, Video resolution enhancement by using discrete and stationary wavelet transforms with illumination compensation, Signal, Image and Video Processing, vol.9, no.1, pp.87-92, 2015.
15. J. Lee, R. Gutierrez-Osuna and S. S. Young, Silk: Scale-space integrated lucas-kanade image registration for super-resolution from video, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.2282-2286, 2013.
16. C. Dong, C. C. Loy, K. He and X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Analysis and Machine Intelligence, vol.38, no.2, pp.295-307, 2016.
17. D. Keren, S. Peleg and R. Brada, Image sequence enhancement using sub-pixel displacements, Proc. of the Computer Society Conference on Computer Vision and Pattern Recognition, pp.742-746, 1988.
18. L. Lucchese and G. M. Cortelazzo, A noise-robust frequency domain technique for estimating planar roto-translations, IEEE Trans. Signal Processing, vol.48, no.6, pp.1769-1786, 2000.
19. B. Marcel, M. Briot and R. Murrieta, “A Frequency Domain Technique Based on Energy Radial Projections for Robust estimation of Global 2D Affine Transformation”, Computer Vision and Image Understanding , vol.81, Issue 1, pp.72-116, 2001.
20. P. Vandewalle, S. Susstrunk and M. Vetterli, “A frequency domain approach to registration of aliased images with application to super-resolution”, EURASIP Journal on Applied Signal Processing, vol.2006, p.233, 2006.
21. Zhi-Song Liu, Wan-chi Siu and Yui-lam Chan, “Photo-Relastic Image Super-Resolution via Variational Autoencoders”, IEEE Transactions on Circuits and Systems for video Technology, 2020.
22. Gholamreza Anbarjafari, “”Video resolution Enhancement Using Deep Neural Networks And Intensity Based Registrations”, International Journal of Innovative and Control Volume 14, Number 5, October 2018.
23. Xiph.org Test Media, http://media.xiph.org/video/derf/, 2016