ISSN: 0976-9102 (ONLINE)
DOI: 10.21917/ijivp.2017.0218

REAL-TIME VIDEO SCALING BASED ON CONVOLUTION NEURAL NETWORK ARCHITECTURE

Safinaz S[1] and Ravi Kumar A.V[2]

[1]Department of Electronics and Communication Engineering, Sir M. Visvesvaraya Institute of Technology, India
[2]Department of Electronics and Communication Engineering, SJB Institute of Technology, India

Abstract

In recent years, video super resolution techniques becomes mandatory requirements to get high resolution videos. Many super resolution techniques researched but still video super resolution or scaling is a vital challenge. In this paper, we have presented a real-time video scaling based on convolution neural network architecture to eliminate the blurriness in the images and video frames and to provide better reconstruction quality while scaling of large datasets from lower resolution frames to high resolution frames. We compare our outcomes with multiple exiting algorithms. Our extensive results of proposed technique RemCNN (Reconstruction error minimization Convolution Neural Network) shows that our model outperforms the existing technologies such as bicubic, bilinear, MCResNet and provide better reconstructed motioning images and video frames. The experimental results shows that our average PSNR result is 47.80474 considering upscale-2, 41.70209 for upscale-3 and 36.24503 for upscale-4 for Myanmar dataset which is very high in contrast to other existing techniques. This results proves our proposed model real-time video scaling based on convolution neural network architecture's high efficiency and better performance.

Keywords:

Image Scaling, Convolution Neural, Network, Super Resolution

1. INTRODUCTION

The In recent years, high definition devices such as HDTV (High-definition television), Smart phones, LAPTOPS, iPad, MacBook Pro and UHDTV (Ultra-high-definition television) have gained immense popularity due to its high resolution quality. Therefore there is an extensive demand of super-resolution in this modern era.

Therefore in recent years Super resolution becomes one of most vital technique for video editing and post-processing applications. Super-resolution is a technique of enhancing the low resolution images or video frames into high resolution frames and images. Super Resolution approach uses neighboring pixels to recover the lost pixels and provide better quality [10]. In many applications such as medical [1], satellite imaging [2], surveillance [3], HDTV [4], video coding or decoding [28]–[30], stereoscopic video processing [31], [32] and face recognition [5] the use of super resolution becomes mandatory requirement.

Super resolution approach is use to extract high-frequency information from the images and video frames with low resolution quality to reconstruct the original image by eliminating the ringing effect [11]. Hence, Super resolution technique needs high amount of accuracy and speed for the processing of video frame sequences and images. Earlier techniques such as Lanczos, bilinear, and bi-spline provides poor quality of images and video frames with number of visual artifacts like ring, blocking and blurring. However, they are cost efficient and can be easily implemented on chip. As a result of poor resolution they cannot provide required precise high quality of images and video frames. Many issues occurs in the hardware implementation of video scaling such as high computational complexity, large memory requirements, requirement of high resolution quality video, pixel replication and redundancy in pixels. Therefore, this motivates us to implement our video scaling model via software.

Therefore, to eliminate these drawback, in recent years a high resolution CNN (Convolution Neural Network) technique [6], [7] come in the existence. The most dynamic advantage of CNN is that it can easily train with large datasets such as ImageNet [8] and Myanmar dataset [9] by using parallel computing on GPU. These datasets are very bulky in size which can be a challenging aspect for other existing techniques. CNN techniques are much faster than the conventional techniques due to its easy training and pure feed-forward methods. However, still most of the existing techniques cannot reconstruct the video frames as efficiently as required. Therefore, in this paper, we present a real-time video scaling approach based on RemCNN technique to provide high resolution scaling for images and video frames.

Our proposed technique RemCNN provides better efficiency and performance in contrast to the existing approaches by eliminating blurriness in the images or video frames to recover the original images and its information. In practical, the key reason of noise occurrence is the difference between the training samples of the training datasets and testing samples of actual application scenes. Therefore, to eliminate this type of noise the proper classification of actual application scenes through Super Resolution (Scaling) approach is necessary so that training samples becomes more similar to that of actual content [12]. In recent years many high resolution devices such as TV (Televisions), laptops and mobile phones developed. However, still many issues such as bulk storage, poor quality and transmission overhead faced by the subscribers. Therefore, to counter these type of conventional issues our proposed RemCNN can be prove very vital technique to help researchers and industries considering the current scenarios.

Video Scaling techniques can be partitioned into two parts such as multi-frame and single-frame based approaches [33], [14]. Single image based approach mostly utilizes interpolation or example techniques due to their least computational cost. However, in this single-frame based approach resources becomes limited which reduces the system performance hence image quality. Therefore, Video Scaling with multi-frame based approach becomes an import aspect for current scenarios in real time to get better quality reconstructed image. Multi-frame based Video Scaling consists of either reconstruction approach or example based approach or combination of them. However, reconstruction approach provides better fidelity but cannot handle large datasets and large motions. On the other hand, Example-based approaches comes with better performance but mostly depends on quality training [34-35]. Therefore to counter these problems our proposed video scaling approach is highly capable which rely upon RemCNN (Reconstruction error minimization Convolution Neural Network) architecture. In this proposed model we use sparse coding reconstruction technique to eliminate the error which generated after feature extraction. We use SReLU (Sparse Rectified Linear Unit) to describe non-linearity. Sparse Coding Based Architecture (SCA) considered to provide better complex relationship between input low resolution images and its generated output high resolution images

However, previous studies [36-37] consists of some limitations related to its high resolution and image reconstruction when upscaling factor increases. Previous experimental results demonstrates that the efficiency of a system drastically decreases whenever upscaling factor increase. This is due to high frequency component of an image is difficult to extract when scaling factor increases as noise and blurriness level also increases. Therefore our model concentrate on maintaining the efficiency of a system even if upscaling factor further increases. However, to provide better efficiency we apply parallel computing on GPU using CAFFE framework regardless of its upscaling factor. Our experimental results demonstrates that the performance of our video scaling model with RemCNN architecture outperforms the existing techniques in terms of scaling factor enhancement outcomes, quality high resolution, noise elimination and precise image reconstruction.

This paper is organize in following sections which are as follows. In section 2, we describe about the video scaling issues and how they can be eliminate by our proposed model. In section 3, we described our proposed methodology. In section 4, experimental results and evaluation shown and section 5 concludes our paper.

2. VIDEO SCALING ISSUES

There are many types of issues which can occur while scaling (either upscaling or downscaling) of images and video frames. In [13], Image Fusion and Super-Resolution with Convolutional Neural Network adopted to eliminate blurriness and provide sharp images for digital photography. In this process Zhong et al. [13] faces pixel level image fusion issues. In [18], 3D Video Super-Resolution Using Fully Convolutional Neural Networks has been proposed to sort out redundancy, degradation in quality of fused image and huge data size problems. In [16], Video Super-Resolution with Convolutional Neural Networks adopted to eliminate the problems of video super-resolution. This paper [16] consist problem of ill posed in reconstruction of high dimension super resolution image and training of large datasets is also a vital issue.

In [14] Image super-resolution: The techniques, Applications, and future provided to review the recent super resolution works and its applications.In that process, the biggest challenge faced by Linwei Yue is that to maintain the quality of resolution in motioning conditions. In [19], A Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network applied. This paper experiences global ill-posed reconstruction SR problem and Single Image Super Resolution (SISR) problem which increases the computational complexity of the model. In [17], Image Super-Resolution Based on Convolution Neural Networks Using Multi-Channel Input proposed to get better feature extraction to reconstruct the image. The problems faced by Youm et al. [17] achieves this objective using gradient exploding and vanishing, and ill posed super resolution problem. In [15], high-quality video/image super-resolution accelerated using GPU to get better performance. This paper has experienced the challenge of running speed requirement for 4K video processing. In [38], Visualizing and Understanding Convolutional Neural Networks proposed to get better feature extraction and enhancement in image quality. There are two challenges such as training of large datasets like ImageNet dataset with elimination of error and poor capturing of pixels by higher layers are widely faced in [38]. Bayesian Adaptive Video Super Resolution model presented to get better high resolution reconstructed image with great feature extractions and performance degradation issues occurs whenever scaling factor increases. In [40], learning a Mixture of Deep Networks for Single Image Super-Resolution model proposed which contains ill-posed, complex mapping of low-resolution images and inverse image recovery problems.

In our paper, the proposed model compared with many existing Super Resolution Video scaling approaches based on CNN framework and there are multiple stages such as shrinking, mapping with sparse coding last layer on which the Convolution Neural Network (CNN) framework rely upon. This stages helps to eliminate the above mentioned poor quality and global ill-posed image reconstruction issues in existing approaches. Dataset videos such as Myanmar video tested with our model and the testing outcomes describes that it can quickly reconstruct the precise information of the video datasets.

The performance of the Super Resolution video scaling architecture significantly increases by using CNN framework. To further improve the performance of the model and speed up the large datasets GPU computing used on a CAFFE framework. CAFFE framework not only accelerate the speed of large datasets but also increases the reconstruction quality of images and video frames. The performance of the system remains same in our system regardless of upscaling factor due to fast parallel computing and sparse coding reconstruction architecture. Sparse coding reconstruction technique helps to eliminate sufficient amount noise in image pixels and ill posed problem and reconstruct an efficient original high resolution image.

3. PROPOSED METHODOLOGY

In this section, it is explained the results of research and at the same time is given the comprehensive discussion. Results can be presented in figures, graphs, tables and others that make the reader understand easily [2], [5]. The discussion can be made in several sub-chapters.

3.1 VIDEO SCALING USING CNN ARCHITECTURE

In recent years, Convolution neural networks (CNN) gains extreme popularity due to its large success in the field of image or video scaling and image classification [20] [21]. CNN can also be easily applied in the fields of face detection [24], pedestrian recognition [25] and object detection [22] [23]. CNN provides fast computation for large training database such as ImageNet [8], Myanmar [9] and videoset4 [39]. There are multiple factors which makes CNN architecture efficient and helps in enhancing the performance of the system [16] [39].

- It helps in the implementation of the training datasets on the efficient and powerful GPU [21] framework such as CAFFE.
- It uses ReLU (Rectifier Linear Unit) [26] to provide better performance and fastening speed in training and testing of datasets.
- It can easily train large datasets like Myanmar datasets [9].

3.2 IMAGE RECONSTRUCTION ARCHITECTURE

In recent years, precise image reconstruction from low-level resolution to high-level resolution image becomes a mandatory requirement. In previous work many techniques or approaches are applied to reconstruct a better quality image. However very few technique are able to provide required high resolution reconstructed image. One technique, which shown high accuracy outcomes and better PSNR performance for image reconstruction, is RemCNN (Reconstruction error minimization Convolution Neural Networks). In this paper, to compute large training datasets with ultra-high speed, GPU computing used in CAFFE framework. To make our system more precise and eliminate sufficient amount of noise from the image or video frame we apply here sparse coding reconstruction technique for a CNN architecture.

The architectural viewpoint for sparse coding reconstruction method is given in Fig.1 which shows the architecture diagram of reconstruction of image. Consider a single low-dimension video frame. In our proposed model patch based feature extracted for each frame in a video. Then all the frames are down-sampled to the intermediate frames. Then for each frame mapping is require. Then frames are up sampled to the desired size. The difference of up sample and down sample frames fed to sparse coding image reconstruction block to reconstruct image to the original quality. Our proposed model outperforms existing techniques by eliminating the error present in the up sampled image and down sampled image. The reconstruction of image or video frames consists of total five stages in our proposed model.

There are multiple stages in our proposed model image reconstruction architecture which are as follows.

Patch based feature Extraction: Our proposed RemCNN technique first performs patch based feature extraction on each original video frame without interpolation. We represents our input image as X_S.Our input image X_S convoluted to a group of filters to get high dimensional feature vector for each frame. In our model group of filters consists of multiple parameters such as S_1, F_1, M_1. As our model perform feature extraction directly on the original frames, the filter size of first sheet S_1 can be as $S_1 = 5$.The number of channels we have adopted here out of YC_bC_r is only channel Y hence the number of channel $M_1 = 1$. Here, $F_1$is the number of feature dimension, which has to determine. This feature dimension of first layer can be presented as *Conv* (5,*s*,1). Here, *s* can be represented as the first sensitive variable.

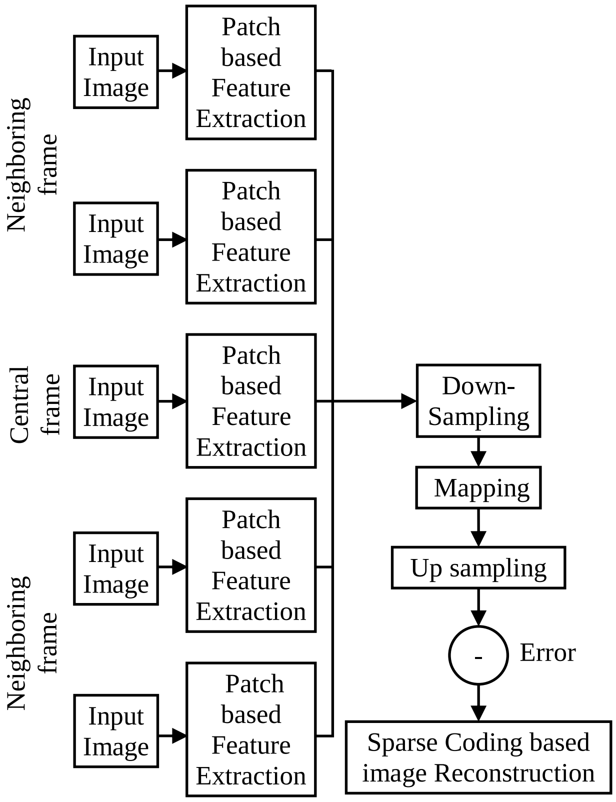

Neighboring frame
Input Image
Patch based Feature Extraction
Input Image
Patch based Feature Extraction
Central frame
Input Image
Patch based Feature Extraction
Neighboring frame
Input Image
Patch based Feature Extraction
Input Image
Patch based Feature Extraction
Down-Sampling
Mapping
Up sampling
-
Error
Sparse Coding based image Reconstruction

Fig.1. Sparse Coding based Image Reconstruction Architecture

Down-Sampling: In existing techniques after feature extraction directly mapping presented. Then high dimension features converted into the high resolution features. This increases computational overhead of the system and degrades the performance due to large size of *s*.

Therefore to eliminate this drawback of the existing techniques here we first down sample the features extracted by the video frames. This approach can also be observed in high-level vision methods to decrease computational cost.

For the same concern we have down sampled the features of all the layers to decrease the feature dimension *s*.The filter size for second layer considered as $S_2 = 1$ to perform linearly with features. The feature dimension for second layer can be presented as $F_2 = g \ll s$. Now feature dimensions are decreased from *s* to *g*. Here, *g* is the second sensitive variable which calculates the amount of downsampling. This feature dimension of second layer (1×1) can be presented as *Conv* (1,*g*,*s*).This technique reduces large amount feature dimension.

Mapping: It is the most vital phase of this proposed algorithm which enhances the performance of the model. This is a non-linear type mapping. In mapping, width and depth are two factors which are most affected. Here, width represent the number of filters present in a layer and depth represents the total number of layers. This operation perform non-linear mapping on each high-dimensional feature. In existing techniques mapping experiments not implemented on large deep networks which helps us to create a more significant non-linear mapping layer. To achieve this we consider a medium filter of size $S_3 = 3$. Then, to provide better efficiency we utilize multiple 3×3 layers. The complexity and accuracy performance calculated by a sensitive variable *d*. Each mapping layer consists of similar number of filters $F_3 = g$. This non-linear mapping can be presented as *Conv*(3,*g*,*g*).

Up-sampling: It is the reverse procedure of the down-sampling. To decrease the feature dimensions down-sampling used which helps in the reduction of computational complexities and produces a high quality video frame or image. Therefore, to generate a high quality image after mapping an up-sampling layer introduced. To retain synchronization between both the layers down-sampling and up-sampling we implemented 1×1 layers. As it is an inverse of down-sampling, the up-sampling layer can represented as *Conv*(1,*s*,*g*). This layer increases the performance of the system.

Sparse coding based image reconstruction: The final part of the image reconstruction is sparse coding based image reconstruction which used to reform a high quality image by eliminating the error produced in up-sampling and down-sampling. Then the outcome (weight parameter) is directly a reformed image with high quality. Here we have taken 9×9 filter layers and the sparse coding layer can be presented as *SparseCode*(9,1,*s*).

3.3 SPARSE RECTIFIED LINEAR UNIT (SReLU)

After each layer, ReLU (Rectified Linear Unit) used for the activation function. In our model we have used Sparse Rectified Linear Unit (SReLU) instead of conventional ReLU. The activation function for SReLU can be,

$$f\,(y_j)=\max(y_j,0) + b_j \ \min(0,y_j) \tag{1}$$

Here, y_j is the input for the activation function *f*, *j* represents the channel and b_j represents the coefficient of negative phase. In existing techniques b_j kept as zero but in for SReLU technique b_j is user-defined. SReLU is a key to eliminate the dead features [27] generates in ReLU by zero gradient vectors. This helps to test parameters of multiple networks for different designs to its full capacity. Our experimental outcomes demonstrate that the SReLU networks is comparatively more efficient and stable. This method increases accuracy and speed as well.

3.4 MODELLING TO REDUCE COMPUTATIONAL COMPLEXITY AND COST FUNCTION

3.4.1 Computational Complexity:

In existing techniques the computational complexity remains very high which degrades the overall performance of the system. The reason for high computational complexity and cost function is the use of conventional *ReLU* and drawback in the design architecture. In existing approaches computational complexity can be calculated as:

$$O\left\{\left(S_1^2F_1+F_1S_2^2F_2+F_2S_3^2\right)S_{hr}\right\} \tag{2}$$

Our proposed model consists of very low computational complexity. This is due to its efficient and accurate modern design architecture and use of sparse rectified

$$O\ \{(25s + gs + 9dg^2 + sg + 81s)\ S_{lr}\} = O\{(9dg^2)\} \tag{3}$$

Linear unit (SReLU) which helps in increasing the speed and avoiding time lapse by eliminating the dead features. In our proposed model computational complexity calculated as:

3.4.2 Cost Function:

In our model cost function described in terms of MSE (Mean Square Root) function. The following equation represent the cost function which used in previous techniques:

$$\min_{\phi}\frac{1}{F}\sum_{k=1}^{F}\left\|F\left(A_g^k;\varphi\right)-B^k\right\|_2^2 \tag{4}$$

Here, A_g^k and B^k are the k^{th} low and high resolution image pair in training. φ is the parameter of the output system function $F\left(A_g^k;\varphi\right)$. The efficiency of these parameters are maintained by utilizing standard back propagation approach with stochastic gradient.

3.4.3 Sparse Coding Reconstruction:

Network Architecture: Sparse Coding Based Architecture (SCA) considered to provide better complex relationship between input low resolution images and its generated output high resolution images. This architecture provides better performance and increases high amount of accuracy. This SCA (Sparse Coding Based Architecture) implemented in corporation with neural networks to reconstruct a high resolution image from the original low-resolution image using LIST (Learned Iterative Shrinkage and Thresholding) approach [40]. The Fig.2 shows the architectural diagram of Sparse Coding Based Architecture (SCA).

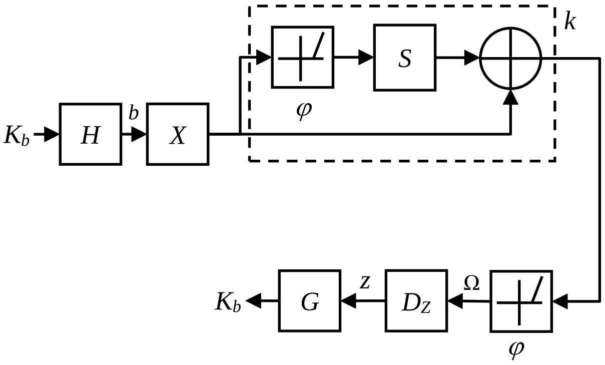


Fig.2. Architectural Diagram of Sparse Coding Based Architecture (SCA)

The objective of our model is to provide high amount of selectivity for the output high resolution image frame by applying SCA based LIST approach on each input frame in coordination with neural networks. We use SReLU (Sparse Rectified Linear Unit) to describe non-linearity. The use of Reconstruction error minimizer Convolution Neural Networks (RemCNN) with SCA reduces high amount of computational cost and enhances efficiency with a large extent. For each output high resolution frame a weight map generated based on pixels. Each generated weight-map multiplied with its equivalent pixels for every output frame. Then all the products of frames are summed up to reconstruct an original image frame. The reconstructed original image $F(b;\ominus)$ can be represented as:

$$F\left(b;\Theta\right)=\sum_{k}^{M}X_k\left(b;\varphi_{\Omega}\right)\odot F_{C_k}\left(b;\varphi_{C_k}\right) \tag{5}$$

Here, *b* represent the input low-resolution image, function $X_k\left(b;\varphi_{\Omega}\right)$ denotes the behavior of generated weight maps and another function $F_{C_k}\left(b;\varphi_{C_k}\right)$ denotes the output high resolution frame C_k. The Eq.(5) represents point wise multiplication of weighted map with its pixels for each reconstructed output image.

The Eq.(6) the loss elimination between the input low-resolution frame and the estimated output frame in training,

$$\min_{\Theta}\sum_{l}\left\|F\left(b_l;\Theta\right)-z_l\right\|_2^2 \tag{6}$$

Here, function $F\left(b_l;\Theta\right)$ denotes our output, b_l is the l^{th} high resolution frame and z_l represents the corresponding low-resolution image. Θ denotes a group of all parameters of our model. The combination of Eq.(5) and Eq.(6) provides the cost function of our proposed model.

$$\min_{\varphi_\Omega\left\{\varphi_{C_k}\right\}_{k=1}^{M}}\sum_{l}\left\|\sum_{k}^{M}X_k\left(b_l;\varphi_\Omega\right)\odot F_{C_k}\left(b_l;\varphi_{C_k}\right)-z_l\right\|_2^2 \tag{7}$$

4. RESULTS AND ANALYSIS

We compute our outcomes with the similar dataset (Myanmar) as used in [9] to compare the performance and efficiency of our model to the existing techniques discussed in the related work. Our model is trained on different large dataset like Myanmar [9]. Testing results shows that our model outperforms most of the existing techniques in terms of PSNR and reconstruction efficiency. We have tested our model for different up scaling factors (2, 3 and 4).

Our results shows accuracy and reconstruction efficiency increment to a large extent. Our model needs less amount of execution time to provide effective video scaling. Our model implemented on 64-bit windows 10 OS with 16GB RAM which consists on INTEL (R) core (TM) i5-4460 processor. It consists of 3.20GHz CPU. We have compared our model with Enhancer [41], Draft-CNN [43], Bayesian [39] and Bayesian-MB [42] and many other existing techniques.

4.1 IMPLEMENTATION DETAILS

We have implemented our extensive experiments on large 4K video Myanmar dataset. In modern era, the availability of 4K monitors is highly increased. Therefore, there is a huge demand of low resolution videos to high-resolution videos in market. These high resolution video can be achieve through upscaling factor. Therefore we have used different upscaling factors to achieve these objectives by measuring performance and accuracy of the model for upscaling factor 2, 3 and 4. Myanmar dataset contains total 57 scenes. In these dataset, 50 scenes used for training and 7 scenes for testing for different up-scaling factors. All the experiments are undertaken on the MATLAB 16b framework in configuration with CAFFE.

4.2 COMPARATIVE STUDY

Here, we have taken 7 scenes for testing out of 57 total scenes in Myanmar dataset video considering upscale-2 as used in [49]. All the scenes are compared to nine most popular existing techniques.Scene-2 represent the famous Myanmar temple which consists of total 594 frames. Our proposed technique RemCNN gives 48.0492 dB PSNR. Scene-8 represents Myanmar golden temple which consists of 354 frames. Our proposed technique RemCNN gives 36.99 dB PSNR for scene-8 which is little less compare to other existing technique. It is an exceptional case in our proposed model. Scene 18 and 33 represents snake and Buddha temple scenes in Myanmar video. Both the scenes consists of 632 frames and for both scenes our proposed technique RemCNN gives highest PSNR as 52.274 and 53.198 dB. Scene 25 and 45 represents yoga scene by a man and horse scenes in Myanmar video. Both the scenes consists of 594 frames and for both scenes our proposed technique RemCNN gives PSNR as 47.031and 49.671 dB.Scene 48 represent tiger scene in Myanmar video. Our proposed technique RemCNN gives PSNR as 47.42 dB for this scene. Similarly, same scenes are used to compute PSNR considering upscale-3 and upscale-4.Our proposed technique shows highest PSNR for scene-1 (temple) which consists of 816 frames. The PSNR results for different upscaling factor 2, 3 and 4 are 54.07, 48.96 and 45.05 dB which is much better than the existing techniques. The percentage improvement of our proposed model in contrast to other conventional techniques is very high.Scene-48 gives highest improvement of 22.17% considering upscale-2. Similarly, scene-45, 33 and 18 gives improvement of 10.21%, 16.37% and 15.22% respectively. However, scene-8 and 25 gives little less accuracy considering upscale -2.Our model gives best improvement result for scene-1 as 28.83% considering upscale-2.

Similarly, Scene-18 gives highest improvement of 15.22% considering upscale-4.Similarly, scene-2, 8, 45 and 48 gives improvement of 3.95%, 12.46%, 2.18% and 14.07 % respectively. However, scene-25 and 33 gives little less accuracy considering upscale-4. Our model gives best improvement result for scene-1 as 28.17% considering upscale-4. Similarly, our model gives best improvement result for scene-1 as 28.15% considering upscale-3.Average PSNR improvement considering upscale-2 and upscale-4 is 7.79% and 4.45%.

The Table.1 shows the comparison of different scenes of a Myanmar dataset for multiple existing techniques. In Table.2 comparision Upscale 4 is given. The following results shows that our average PSNR result is 47.80474 dB considering upscale-2, 41.70209 dB for upscale-3 and 38.24503 dB for upscale-4 (Table.3) considering all seven testing scenes which is much better than the existing techniques for Myanmar dataset. Similarly, Table.4 and Table.5 represent SSIM (structural similarity index) comparison with recent existing techniques considering upscale-2 and 4 for scene 2,8,18,25,33,45 and 48. The Table.7 represent SSIM comparison with VSRnet and MCResNet considering upscale 2, 3 and 4 for scene-1 which is better than existing techniques.

Table.1. PSNR Values (in dB) of the SR Frame for Various Methods and Test Scenes (best results are shown in Bold) considering upscale-2

Scenes	**Scene-2**	**Scene-8**	**Scene-18**	**Scene-25**	**Scene-33**	**Scene-45**	**Scene-48**
bicubic	45.27	38.18	41.43	44.4	40.22	42.43	33.9
bi-level	46.12	39.94	43.04	46.69	42.95	43.72	36.1
SDMF-B	46.81	40.08	43.41	47.52	43.08	44.07	36.2
SDMF-R	46.41	40.32	43.69	47.68	43.55	44.18	35.66
MDSF	46.79	40.34	43.37	47.37	43.27	44.05	36.55
MDMF-B	47.66	40.59	43.92	48.45	43.68	44.49	36.67
MDMF-R	46.86	40.6	44.19	47.83	44.05	44.28	36.07
MDMF-B-VT	48.14	40.98	44.32	49.19	44.49	44.6	36.91
MDMF-R-VT	47.41	41.05	44.46	48.59	44.48	44.62	36.64
Proposed	**48.0492**	**36.99**	**52.274**	**47.031**	**53.198**	**49.671**	**47.42**

Table.2. PSNR values (in dB) of the SR frame for various methods and test scenes (best results are shown in bold) consideing upscale-4

Scenes	**Scene-2**	**Scene-8**	**Scene-18**	**Scene-25**	**Scene-33**	**Scene-45**	**Scene-48**	**Average**
Bicubic	39.58	32.13	35.65	36.1	32.15	36.13	27.25	34.14
bi-level [44]	40.5	32.46	36.37	37.02	33.44	36.71	28.03	34.93
NE+NNLS [45]	41.32	33	36.76	37.9	33.79	37.12	28.04	35.42
NE+LLE [46]	41.12	32.95	36.82	37.78	33.94	37.27	28.2	35.44
ANR [47]	41.32	32.81	36.76	37.49	34	37.35	28.26	35.43
SR-CNN [48]	43.17	33.4	37.5	38.35	34.57	37.9	28.73	36.23
Enhancer [41]	40.62	32.09	36.44	37.44	34.67	37.15	27.75	35.17
Bayesian [39]	39.18	31.73	35.7	35.34	32.14	35.76	26.76	33.8
MDMF-B-VT [49]	43.48	33.48	37.68	39.03	34.92	38.42	28.75	36.54
MDMF-R-VT [49]	42.9	33.42	37.65	38.75	34.86	38.1	28.49	36.31
Proposed	45.2676	38.2453	44.44899	38.96423	28.45522	38.87859	33.45522	38.24503

Table.3. PSNR Values (in dB) of the SR frame for various methods and test scenes considering Upscale-2,3,4 for Myanmar dataset

Scenes	**Proposed (upscaling -2)**	**Proposed (upscaling-3)**	**Our Proposed (upscaling-4)**
Scene-2	48.0492	49.34129143	45.26765851
Scene-8	36.9900	41.61521459	38.24530073
Scene-18	52.2740	48.24806934	44.44898805
Scene-25	47.0310	42.87746065	38.96423093
Scene-33	53.1980	31.56951803	28.45521909
Scene-45	49.6710	42.19030456	38.87859171
Scene-48	47.4200	36.07273785	33.45522214
Average	**47.80474**	**41.70209**	**38.24503**

Table.4. PSNR comparison for upscale 2, 3 and 4 with MD MFB-VT and MD MFR-VT for scene-1

	MD MFB-VT	**MD MFR-VT**	**RemCNN**
Upscale-2	38.48	40.04	54.07
Upscale-3	34.42	35.18	48.96
Upscale-4	31.85	32.36	45.05

Table.5. SSIM Values (in dB) of the SR frame for various methods and test scenes (best results are shown in bold) consideing upscale-2

Scenes	**Scene-2**	**Scene-8**	**Scene -18**	**Scene-25**	**Scene-33**	**Scene-45**	**Scene-48**	**Average**
Bicubic	0.983	0.9738	0.9738	0.9917	0.9786	0.9718	0.9668	0.9771
bi-level [44]	0.9879	0.9842	0.9849	0.9961	0.9904	0.981	0.9808	0.9865
NE+NNLS [45]	0.9851	0.9824	0.982	0.9938	0.9879	0.9772	0.9774	0.9837
NE+LLE [46]	0.9834	0.9817	0.9816	0.9936	0.9889	0.9776	0.9785	0.9836
ANR [47]	0.9857	0.9832	0.9833	0.9952	0.9902	0.9791	0.9799	0.9851
SR-CNN [48]	0.9859	0.9852	0.9844	0.9955	0.9907	0.9797	0.9826	0.9863
Enhancer [41]	0.9854	0.9823	0.9844	0.9938	0.9908	0.9764	0.9751	0.984
Bayesian [39]	0.9874	0.9828	0.9842	0.9954	0.99	0.979	0.977	0.9851
MDMF-B-VT [49]	0.9882	0.9884	0.9877	0.997	0.9937	0.9812	0.9846	0.9887
MDMF-R-VT [49]	0.9882	0.9882	0.9884	0.9967	0.9938	0.9823	0.9821	0.9885
Proposed	0.9971	0.9926	0.9963	0.996	0.9801	0.9942	0.9872	0.9919

Table.6. SSIM values (in dB) of the SR frame for various methods and test scenes (best results are shown in bold) consideing upscale-4

Scenes	**Scene-2**	**Scene-8**	**Scene-18**	**Scene-25**	**Scene-33**	**Scene-45**	**Scene-48**	**Average**
Bicubic	0.9648	0.9013	0.9122	0.9515	0.8899	0.9101	0.8514	0.9116
bi-level [44]	0.9662	0.9099	0.9209	0.9546	0.914	0.9155	0.873	0.922
NE+NNLS [45]	0.9691	0.9145	0.9243	0.9622	0.9157	0.9193	0.871	0.9252
NE+LLE [46]	0.9675	0.9187	0.9249	0.9607	0.9188	0.9211	0.8757	0.9268
ANR [47]	0.9691	0.9107	0.9243	0.9587	0.9206	0.9226	0.878	0.9263
SR-CNN [48]	0.9703	0.9198	0.928	0.9633	0.923	0.9253	0.8883	0.9311
Enhancer [41]	0.9695	0.9121	0.9308	0.9621	0.9304	0.9267	0.8679	0.9285
Bayesian [39]	0.966	0.8972	0.9183	0.9473	0.8945	0.9083	0.8393	0.9101
MDMF-B-VT [49]	0.9737	0.9266	0.9331	0.9702	0.9363	0.934	0.8921	0.938
MDMF-R-VT [49]	0.974	0.925	0.9341	0.9687	0.9374	0.9316	0.8842	0.9364
Proposed	0.99	0.9515	0.9837	0.9741	0.8795	0.9617	0.9224	0.9519

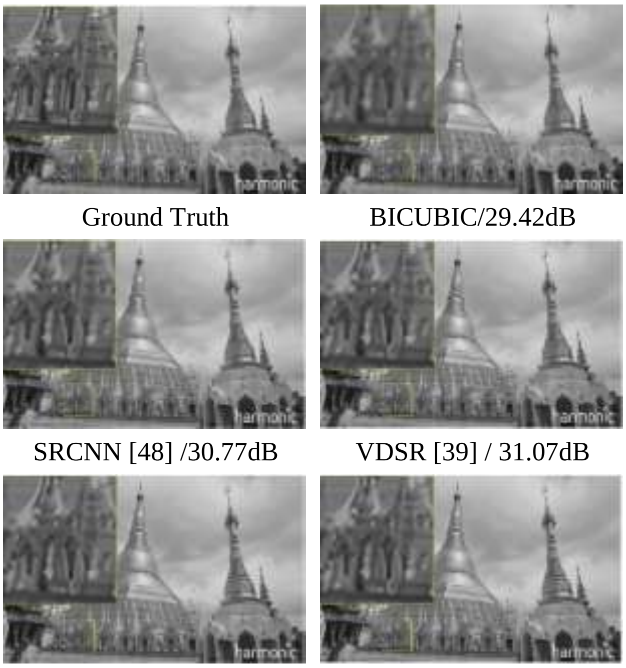
Table.7. SSIM comparison for upscale 2, 3 and 4 with MD MFB-VT and MD MFR-VT for scene-1

	VSRnet	**MCResNET**	**RemCNN**
Upscale-2	0.9679	0.9777	0.9973
Upscale-3	0.9247	0.9387	0.9956
Upscale-4	0.8834	0.8987	99.08

4.3 IMAGE RECONSTRUCTION COMPARISON

Here, we have demonstrated 350th frame of scene-8 as used in all the other existing techniques. The original Myanmar video dataset contains total 57 scenes and its original resolution is 3840×2160. We have shown PSNR and image reconstruction quality comparison with all the conventional techniques. The PSNR result (41.615 dB) outperforms all the existing state-of-the- art techniques. From our experimental results it is clearly visible that our reconstruct frame has better reconstruction quality than any other recent existing techniques.

Table.8. Comparison with different existing techniques



Ground Truth

BICUBIC/29.42dB

SRCNN [48] /30.77dB

VDSR [39] / 31.07dB

DRCN [51] / 31.05dB

VSRnet AMC [16] / 31.02dB

VSRnet MC [16] / 31.12dB

MCResNet [50] / 32.16dB

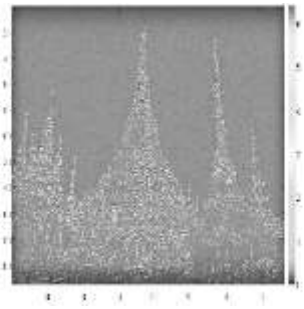
Proposed RemCNN / 41.615dB

4.4 RECONSTRUCTION ERROR MAP

Here, we reconstruct error map for 350th frame of scene-8 shown in Table.9. Reconstruction map is the combination of planning motion and local motion of background and foreground respectively. The reconstruction error depends on the iterations as the iteration increases, the error become decreases. In the final outcomes error become disappears or become negligible using our proposed technique RemCNN.

Table.9 reconstructed error map from original image

Original Image



Reconstructed Error Map

4.5 GRAPHICAL ANALYSIS

The following graphs shows the comparison between our proposed model and existing approaches MD MFB-VT and MD MFR-VT for upscale 2, 3 and 4 considering Myanmar dataset. The Fig.3 shows PSNR comparison considering upscale -2 for the scenes 2, 8, 18, 25, 33, 45 and 48. The Fig.4 shows PSNR comparison considering upscale - 4 for the scenes 2, 8, 18, 25, 33, 45 and 48. The Fig.5 demonstrates PSNR comparison considering upscale-2, 3 and 4 for scene-1 with both recent existing techniques MD MFB-VT and MD MFR-VT. PSNR for upscale-2 using our proposed RemCNN technique is 54.07dB, with upscale-3 is 48.96dB and fro upscale-4 is 45.05dB which is very high compare to other techniques.

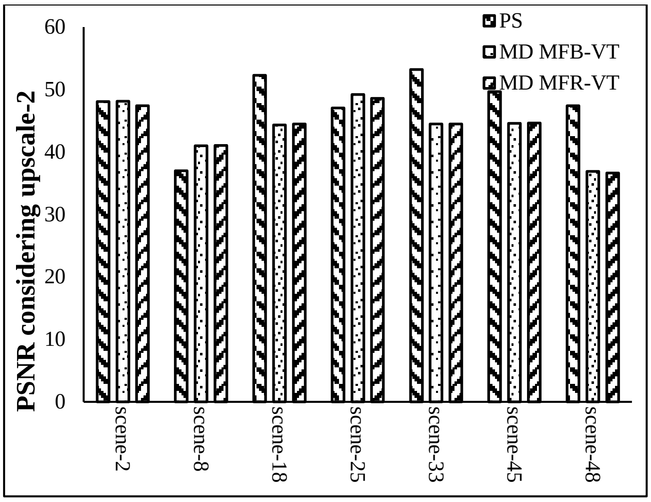


Fig.3. PSNR Comparison for Proposed vs. Existing Techniques for Upscaling Factor 2 for Myanmar Dataset

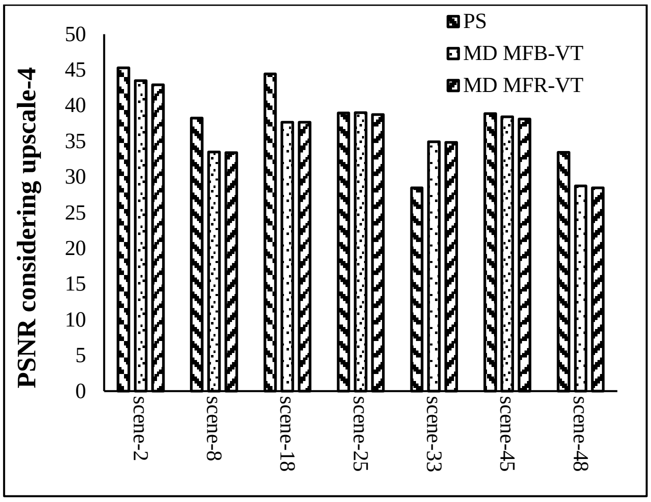


Fig.4. PSNR Comparison for Proposed vs. Existing Techniques for Upscaling Factor 4 for Myanmar Dataset

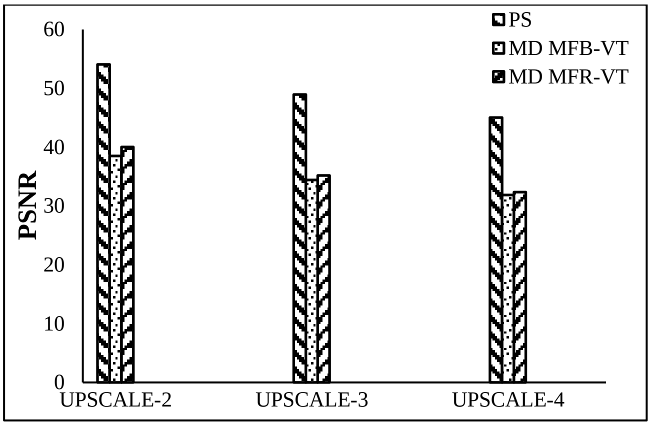


Fig.5. PSNR for Proposed vs. Existing Considering Upscale Factor 2, 3 and 4 for Myanmar Dataset

5. CONCLUSIONS

In current era, huge demand and popularity of high resolution videos made researchers to carry out work in video scaling field to offer ease of accessibility of high-resolution videos to the subscribers. Therefore, we have introduced a real-time video scaling based on convolution neural network architecture to eliminate the blurriness in the images and video frames and to provide better reconstruction quality while scaling of large datasets. CNN architecture helps us to restore high frequency components of the video frames. Our proposed model can easily train the bulky datasets such as Myanmar and Videoset4.Our experimental results shows that our model outperforms many existing techniques in terms of PSNR, fidelity and reconstruction quality. The experimental results shows that our average PSNR result is 47.80474 considering upscale-2, 41.70209 for upscale-3 and 36.24503 for upscale-4 for Myanmar dataset which is very high in contrast to other existing techniques. Our model gives best improvement result for scene-1 as 28.83% considering upscale-2, 28.17% considering upscale-4, 28.15% for upscale-3.

This results proves our proposed model real-time video scaling based on convolution neural network architecture's high efficiency and better performance. Our proposed model can be effectively used in the applications such as medical, satellite imaging, surveillance, HDTV, video coding or decoding, stereoscopic video processing, and face recognition for future purpose to reconstruct efficient images or video frames.

REFERENCES

[1] Wenzhe Shi et.al., "Cardiac Image Super-Resolution with Global Correspondence using Multi-Atlas Patchmatch", *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 9-16, 2013

[2] M.W. Thornton, P.M. Atkinson and D.A. Holland, "Sub-pixel Mapping of Rural Land Cover Objects from Fine Spatial Resolution Satellite Sensor Imagery using Super-Resolution Pixel-Swapping", *International Journal of Remote Sensing*, Vol. 27, No. 3, pp. 473-491, 2006.

[3] L. Zhang, H. Zhang, H. Shen and P. Li, "A Super-Resolution Reconstruction Algorithm for Surveillance Images", *Signal Processing*, Vol. 90, No. 3, pp. 848-859, 2010.

[4] T. Goto, T. Fukuoka, F. Nagashima, S. Hirano and M. Sakurai, "Super-Resolution System for 4K-HDTV", *Proceedings of 22nd International Conference on Pattern Recognition*, pp. 4453-4458, 2014.

[5] B.K. Gunturk, A.U. Batur, Y. Altunbasak, M.H. Hayes and R.M. Mersereau, "Eigenface-Domain Super-Resolution for Face Recognition", *IEEE Transactions on Image Processing*, Vol. 12, No. 5, pp. 597-606, 2003.

[6] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks", *Proceedings of Neural Information Processing Systems*, pp. 1097-1105, 2012.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going Deeper with Convolutions", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2014.

[8] J. Deng, W. Dong, R. Socher and L. Li, "A Large-Scale Hierarchical Image Database", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248-255, 2009.

[9] Harmonic Inc, Avaailable at: http://www.harmonicinc.com/resources/videos/4kvideo-clip-center, Accessed on 2014.

[10] Weisheng Dong, Lei Zhang, Guangming Shi and Xiaolin Wu, "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization", *IEEE Transactions on Image Processing*, Vol. 20, No. 7, pp. 1838-1857, 2011.

[11] Marshall F. Tappen, Bryan C. Russell and William T. Freeman, "Exploiting the Sparse Derivative Prior for Super-Resolution and Image Demosaicing", *Proceedings of IEEE Workshop on Statistical and Computational Theories of Vision*, pp. 1-28, 2003.

[12] L. Zhang et al., "FSIM: A Feature Similarity Index for Image Quality Assessment", *IEEE Transactions on Image Processing*, Vol. 20, No. 8, pp. 2378-2386, 2011.

[13] J. Zhong, B. Yang, Y. Li, F. Zhong and Z. Chen, "Image Fusion and Super-Resolution with Convolutional Neural Network", *Proceedings of Chinese Conference on Pattern Recognition*, pp. 78-88, 2016.

[14] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, L. Zhang, "Image Super-Resolution: The Techniques Applications and Future", *Signal Processing*, Vol. 128, pp. 389-408, 2016

[15] Z. Zhao, L. Song, R. Xie and X. Yang, "GPU Accelerated High-Quality Video/Image Super-Resolution", *Proceedings* of *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, pp. 1-4, 2016.

[16] A. Kappeler, S. Yoo, Q. Dai and A.K. Katsaggelos, "Video Super-Resolution With Convolutional Neural Networks", *IEEE Transactions on Computational Imaging*, Vol. 2, No. 2, pp. 109-122, 2016.

[17] G.Y. Youm, S.H. Bae and M. Kim, "Image Super-Resolution based on Convolution Neural Networks using Multi-Channel Input", *Proceedings of IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop*, pp. 1-5, 2016.

[18] Y. Xie, J. Xiao, T. Tillo, Y. Wei and Y. Zhao, "3D Video Super-Resolution using Fully Convolutional Neural Networks", *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1-6, 2016.

[19] W. Shi et al., "Real-Time Single Image and Video Super-Resolution using an Efficient Sub-Pixel Convolutional Neural Network", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874-1883, 2016.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition", *Proceedings of European Conference on Computer Vision*, pp. 346-361, 2014.

[21] Alex Krizhevsky, Iiya Sutskever and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", *Advances in Neural Information Processing Systems*, pp. 1097-1105, 2012.

[22] W. Ouyang, P. Luo, X. Zeng, S. Qiu, Y. Tian, H. Li, S. Yang, Z. Wang, Y. Xiong, C. Qian, "Deepid-Net: Multi-Stage and Deformable Deep Convolutional Neural Networks for Object Detection", *Proceedings of Computer Vision and Pattern Recognition*, pp. 1-13, 2014.

[23] Wanli Ouyang and Xiaogang Wang, "Joint Deep Learning for Pedestrian Detection", *Proceedings of IEEE*

International Conference on Computer Vision, pp. 2056-2063, 2013.

[24] Yi Sun, Yuheng Chen, Xiaogang Wang and Xiaoou Tang, "Deep Learning Face Representation by Joint Identification-Verification", *Proceedings of Advances in Neural Information Processing Systems*, pp. 1988-1996, 2014.

[25] Christian Szegedy, Scott Reed, Dumitru Erhan, Dragomir Anguelov and Sergey Ioffe, "Scalable, Highquality Object Detection", *Proceedings of IEEE International Conference on Computer Vision*, pp. 1-10, 2014.

[26] V. Nair and G.E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines", *Proceedings of International Conference on Machine Learning*, pp. 807-814, 2010.

[27] Matthew D. Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks", *Proceedings of European Conference on Computer Vision*, pp. 818-833, 2014.

[28] Byung Cheol Song, Shin-Cheol Jeong and Yanglim Choi, "Video Super-Resolution algorithm using Bi-Directional Overlapped Block Motion Compensation and on-the-Fly Dictionary Training", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 21, No. 3, pp. 274-285, 2011.

[29] Edson Mintsu Hung, Ricardo L. de Queiroz, Fernanda Brandi, Karen França de Oliveira and Debargha Mukherjee, "Video Super-Resolution using Codebooks Derived from Key-Frames", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 22, No. 9, pp. 1321-1331, 2012.

[30] Zhengdong Zhang and Vivienne Sze, "Fast: Free Adaptive Super-Resolution via Transfer for Compressed Videos", *Proceedings of Computer Vision and Pattern Recognition*, pp. 1-17, 2016.

[31] Jing Zhang, Yang Cao, Zheng-Jun Zha, Zhigang Zheng, Chang Wen Chen and Zengfu Wang, "A Unified Scheme for Super-Resolution and Depth Estimation from Asymmetric Stereoscopic Video", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 26, No. 3, pp. 479-493, 2016.

[32] Zhi Jin, Tammam Tillo, Chao Yao, Jimin Xiao and Yao Zhao, "Virtual-View-Assisted Video Super-Resolution and Enhancement", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 26, No. 3, pp. 467-478, 2016.

[33] Kamal Nasrollahi and Thomas B. Moeslund, "Super-Resolution: A Comprehensive Survey", *Machine Vision and Applications*, Vol. 25, No. 6, pp. 1423-1468, 2014.

[34] S. Farsiu, M.D. Robinson, M. Elad and P. Milanfar, "Fast and Robust Multiframe Super Resolution", *IEEE Transactions on Image Processing*, Vol. 13, No. 10, pp. 1327-1344, 2004.

[35] M. Protter, M. Elad, H. Takeda and P. Milanfar, "Generalizing the Nonlocal-Means to Super-Resolution Reconstruction", *IEEE Transactions on Image Processing*, Vol. 18, No. 1, pp. 36-51, 2009.

[36] S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them", *IEEE Transaction on Pattern Analysis Machine Intelligence*, Vol. 24, No. 9, pp. 1167-1183, 2002.

[37] Zhouchen Lin and Heung-Yeung Shum, "Fundamental Limits of Reconstruction based Super Resolution Algorithms under Local Translation", *IEEE Transaction on Pattern Analysis Machine Intelligence*, Vol. 26, No. 1, pp. 83-97, 2004.

[38] Matthew D. Zeiler and Rob Fergus, "Visualizing and Understanding Convolutional Networks", *Proceedings of European Conference on Computer Vision*, pp. 818-833, 2014.

[39] W.T. Freeman, T.R. Jones and E.C. Pasztor, "Example-based Super Resolution", *IEEE Computer Graphics and Applications*, Vol. 22, No. 2, pp. 56-65, 2002.

[40] C. Liu and D. Sun, "On Bayesian Adaptive Video Super Resolution", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 2, pp. 346-360, 2014.

[41] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han and Thomas Huang, "Deep Networks for Image Super-Resolution with Sparse Prior", *Proceedings of IEEE International Conference on Computer Vision*, pp. 370-378, 2015.

[42] Video enhancer, Available: http://www.infognition.com/videoenhancer/

[43] Z. Ma, R. Liao, X. Tao, L. Xu, J. Jia, and E. Wu, "Handling Motion Blur In Multi-Frame Super-Resolution", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5224-5232, 2015.

[44] Renjie Liao, Xin Tao, Ruiyu Li, Ziyang Ma and Jiaya Jia, "Video Super-Resolution via Deep Draft-Ensemble Learning", *Proceedings of IEEE International Conference on Computer Vision*, pp. 531-539, 2015.

[45] J. Yang, Z. Wang, Z. Lin, X. Shu, and T. Huang, "Bilevel Sparse Coding for Coupled Feature Spaces", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2360-2367, 2012.

[46] M. Bevilacqua, A. Roumy, C. Guillemot, and M.L. Alberi-Morel, "Lowcomplexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding", *Proceedings of 23rd British Machine Vision Conference*, pp. 13501-13510, 2012.

[47] H. Chang, D.Y. Yeung and Y. Xiong, "Super-Resolution through Neighbor Embedding", *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 1-6, 2004.

[48] R. Timofte, V. De and L. Van Gool, "Anchored Neighborhood Regression for Fast Example-based Super-Resolution", *Proceedings of IEEE International Conference on Computer Vision*, pp. 1920-1927, 2013.

[49] C. Dong, C.C. Loy, K. He and X. Tang, "Learning A Deep Convolutional Network for Image Super-Resolution", *Proceedings of European Conference on Computer Vision*, pp. 184-199, 2014.

[50] Q. Dai, S. Yoo, A. Kappeler and A.K. Katsaggelos, "Sparse Representation-Based Multiple Frame Video Super-Resolution", *IEEE Transactions on Image Processing*, Vol. 26, No. 2, pp. 765-781, 2017.

[51] Dingyi Li and Zengfu Wang, "Video Super-Resolution via Motion Compensation and Deep Residual Learning", *IEEE Transactions on Computational Imaging*, Vol. PP, No. 99, pp. 1-15, 2017.