Align-Filter & Learn Video Super Resolution using Deep learning (AFLVSR)

Padma Reddy AM, Udaya Rani V

Abstract*: *In this article, we focused on the super-resolution technique in computer vision applications. During last decade, image super-resolution techniques have been introduced and adopted widely in various applications. However, increasing demand of high quality multimedia data has lead towards the high resolution data streaming. The conventional techniques which are based on the image super-resolution are not suitable for multi-frame SR. Moreover, the motion estimation, motion compensation, spatial and temporal information extraction are the well-known challenging issues in video super-resolution field. In this work, we address these issues and developed deep learning based novel architecture which performs feature alignment, filtering the image using deep learning and estimates the residual of low-resolution frames to generate the high-resolution frame. The proposed approach is named as Align-Filter & Learn Video Super resolution using Deep learning (AFLVSR). We have conducted and extensive experimental analysis which shows a significant improvement in the performance when compared with the state-of-art video SR techniques.

Index Terms*: *video super resolution, CNN, feature alignment, deep learning.

I. INTRODUCTION

The demand for computer vision-based applications such as object detection, recognition, tracking and medical imaging, etc. has increased drastically. These applications utilize the image or video data to accomplish the desired task. The poor quality of these data such as noisy data, poor illumination, lighting conditions, blurred data and low-resolution data, etc. may lead towards the inappropriate and faulty outcome. A significant amount of researches have been carried out during the last decade to address these issues [1- 4]. Currently, the increasing demand of high-definition display and ultra-high-definition video formats such as 4K (3840x2160) and 8K (7680x4320) has gained attraction from research community and industries to improve the resolution of the data [5]. Moreover, in various applications such as video surveillance, the capturing devices have some limitations as mentioned in [6]. Due to these factors, we get low-resolution (LR) images. However, the current computer vision-based applications urge for high-resolution data. In order to overcome this problem, the super-resolution (SR) technique is considered as a promising solution to generate the high resolution (HR) image from a single image or sequence of LR images. The SR scheme discards the noise, resolution degradation and other components from the data and increases high-frequency components to improve the image resolution [7].

Several methods of super-resolution have been introduced during the last decade. These methods are based on the single-frame SR and multi-frame/video SR. Single image super-resolution has been studied widely and various techniques use the concept of single image SR for video SR [8-9]. The SR is known as an ill-posed problem which requires prior knowledge to obtain the solution by exploiting the spatial correlation of provided LR image. Early attempts of single image SR are not capable to resolve the low-resolution issues because of lack of prior knowledge. Recently, learning-based schemes have drawn attention due to their significant learning process. Recently, deep learning and convolutional neural network (CNN) based methods have emerged as a promising solution for single images SR. These techniques include residual dense network [10], dilated convolutional neural [11], deep convolution neural network [12] and many more. However, in video frames, SR techniques focus on exploiting the temporal and spatial correlations thus it becomes more challenging to perform SR for videos. According to Bare et al. [13], the video super-resolution techniques can be classified as reconstruction-based methods, example-based techniques and hybrid techniques. Reconstruction based SR techniques perform the sub-pixel registration between adjacent frames to construct the high resolution output. These techniques obtain reliable performance for small global motion scenarios but fail to perform for large scales and large motions. Example based methods perform direct mapping between LR and HR frames. Similarly, hybrid methods use both techniques. In recent years, the Convolutional Neural Network (CNN) and optical methods are widely adopted for SR application where motion compensation and ensemble learning techniques can be used to solve the motion and large scale complexities [14-17].

Several studies show the significance of deep learning and CNN in the field of super resolution. Kappeler et al. [18] reported that the CNNs are capable of processing the huge data such as ImageNet because these network models can be trained using GPU accelerated computing. Moreover, after training the CNN, super-resolution process becomes a feed-forward network which is faster than the traditional learning methods. Hence, these techniques are adopted to perform the super-resolution task in video sequences. In [19] authors developed adaptive neural network approach to determine the temporal dependence between frames.

Revised Manuscript Received on November 05, 2019.

Padma Reddy A.M, Research Scholar, REVA University. Professor, Sai Vidya Institute of Tech., Bangalore. Email: padmareddy.am@saividya.ac.in

Dr. Udaya Rani V, School of Computing & Information Technology, REVA University, Bangalore. Email: udayaraniv@reva.edu.in.

Moreover, this network helps to obtain the spatial alignment resulting in robustness against motion. Generally, the traditional deep learning video resolution techniques are based on the motion estimation, compensation and upsampling process such as CNN [18], Spatio-Temporal Networks and Motion Compensation [14,16-17]. However, the performance of these techniques depends on the accurate motion estimation and the generated HR output is obtained by mixing the values of motion compensated LR frames using CNN. Due to this mixture, the CNN may generate the blurry HR frame. Recently techniques [20-22] presented pipeline architecture to address these challenges. The pipeline architecture follows feature extraction, feature alignment, fusion and frame reconstruction. Recently, Wang et al. [23] reported that the feature alignment and fusion are the crucial steps when the video frames have more occlusion, motion and blurriness. Hence, in order to generate the high-quality HR output, we also focus on the feature alignment and fusion. Moreover, during reconstruction, the imperfect alignment and blurriness can result in the poor quality HR image.

In this work, we address the aforementioned issues and introduce CNN based novel architecture for video super-resolution which considers resolution enhancement to generate the HR frame. The proposed approach follows a two-fold scheme where first phase performs feature alignment and second phase performs the feature fusion by considering both temporal and spatial dependence between frames. Below given figure 1 shows sample outcome of proposed approach of video super-resolution.

(a) Low-resolution (b) High Resolution (c) Ground truth

Fig.1. Sample outcome of different super resolution techniques

The main contributions of the proposed approach are as follows:

(a) We introduce CNN based architecture for video super resolution.

(b) A feature alignment module is presented to extract the pyramdical features and align them to mitigate the temporal issues.

(c) Upsampling and filtering network is presented based on the CNN model.

The rest of the article is organized as follows: section II presents brief literature review about recent techniques of super-resolution, section III presents proposed solution for super-resolution, section IV presents comparative experimental analysis and finally, section V presents the conclusion remarks.

II. LITERATURE SURVEY

In this section, we present brief literature review about recent techniques of video super-resolution. As discussed in previous section, the deep learning based schemes have gained huge attraction in computer vision field. The deep learning based techniques are adopted widely in image and video super-resolution due to their significant nature of learning.

According to conventional SR scheme, the batch of LR frames is processed to generate the HR. In this process, each frame is processed multiple times resulting in increased computational complexity. Moreover, the inappropriate temporal analysis may lead towards the degraded performance. In order to overcome these issues, Sajjadi et al. [21] introduced an end-to-end trainable framework which adopts recurrent nature to understand previous frames completely. However, this method uses a single frame processing which leads to generate the suboptimal outcome in case of occlusion. Recently, Haris et al. [22] addressed this issues and developed recurrent back-projection network (RBPN) for SR. In this method, the network processes multiple frames where spatial and temporal information is extracted and fused using recurrent encoder-decoder module. Unlike other schemes, RBPN considers each frame as source of information and process them separately.

Motion compensation and estimation is also considered as an important aspect in SR techniques. The traditional CNN based schemes predicts the motion and align the reference frame. Later, the current LR and reference frames are processed through CNN to generate corresponding HR frame. However, these methods suffer from the issue of accurate motion estimation and complex CNN architecture. In order to address aforementioned issues, Bare et al. [13] focused on real-time video resolution using motion convolutional kernel estimation and introduced a novel approach. According to this approach, the 1D motion convolution kernels are used for motion prediction along with alignment and later gated residual unit (GRU) is used to combine the input and output weights. Caballero et al. [14] presented motion compensation based approach for super resolution.

In super resolution techniques, the spatial and temporal information from each frame play important role to generate the HR frame. 3D-Convolution is known as a promising technique for this purpose but due to increased computational complexity the performance of 3D convolutions may degrade. In order to solve this issue, Li et al. [23] presented spatio-temporal residual network. In this process, spatio-temporal residual block (STRB) is proposed which divides the 3D filters to reduce the dimension. Later, cross-space residual learning is applied to correlate the LR and HR spaces

Wang et al. [5] also developed multi-memory CNN to fully exploit the information between LR frames. This approach performs cascading of optical flow network. The feature extraction process uses residual block to obtain the spatial correlation. Similar to this work, Yi et al. [24], the conventional techniques use single channel and single memory module. Using these techniques, the complete spatial and temporal information cannot be extracted. To address these issues, authors presented multi-temporal ultra-dense memory (MTUDM) network where spatio-temporal correlation features are extracted using an ultra-dense memory block (UDMB). The UDMB is constructed by embedding the convolutional long-short-term memory (ConvLSTM) into ultra-dense residual block (UDRB).

The inaccurate inter-frame information analysis leads towards the poor quality reconstruction of HR frames. In [25], authors presented dense-connected residual network (DCRnet) model to improve the performance of SR. According to this approach, the low frequency components of motion compensated frames are preserved and hierarchal features are extracted from convolutional layers to restore the HR frame.

III. PROPOSED MODEL

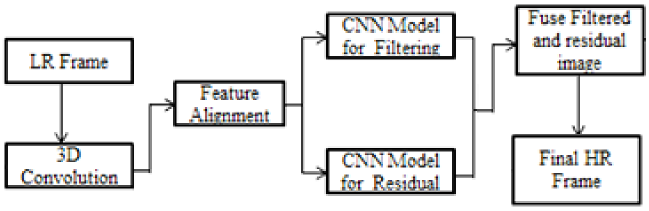
In this section, we present the proposed solution to address the SR related issues using a novel CNN based architecture. First of all, we present feature alignment strategy using a convolutional scheme. In the next phase, we present a novel approach of dynamic upsampling and residual learning. Based on these modules, we design an end-to-end network to solve the SR problems as observed in existing techniques. Below is the list of notations used in this paper for modelling.

Table.1. Notations used in the article

Notation	Meaning
X	LR frames
$\mathcal{G}$	Ground truth frames
N	Temporal radius
X_t	Reference frame
$\hat{Y}$	Output HR frame
L	Total frames
$\mathbb{F}$	Feature
w	Convolution weights
p	offset
K	Convolution location

A. Overview of proposed model

Let us consider that we have $L = 2N + 1$ consecutive low-resolution frames denoted as $X_{[t-N:t+N]}$. The middle frame X_t is considered as reference frame and other two frames are considered as neighboring frames. In this work, our aim is to estimate the high resolution frames from given LR frames. The considered LR frames are down-sampled by a time step t as X_t corresponding to ground truth frames $\mathcal{G}_t$. Table 1 show notations used in this article. Figure 1 presents a graphical overview of proposed AFLSVR technique.



According to the proposed AFLSVR, the low resolution frames are given as input and HR frames are generated as the output of system. As discussed before, feature alignment is an important task to achieve the HR frames hence we present a cascaded convolutional model to generate the aligned frames. Later, we present dynamic upsampling to generate a filtering network and finally, a deep learning based residual learning model is introduced to generate the HR frames. In order to generate the HR using proposed AFLVSR network $\mathbb{N}$ and network parameters α, the SR problem can be defined as:

$$\hat{Y}_t = \mathbb{N}_\alpha(X_{t-N:t+N}) \quad (1)$$

where N denotes the temporal radius. The input to AFLVSR network is given as $W \times H \times L \times C$, where W denotes width, H denotes height, L denotes total consecutive LR frames and C denote the number of channels in the frame.

B. Dynamic Upsampling and filtering network

In this section, we introduce a dynamic filtering and upsampling network to improve the exploitation of spatio-temporal information in given LR frames. Figure 2 shows process of dynamic upsampling and filtering.

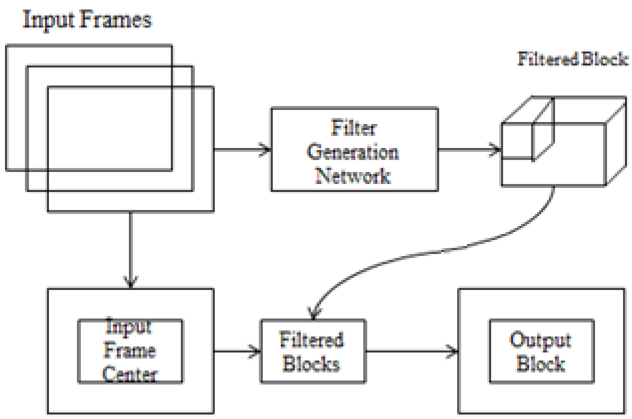


Fig.2. Dynamic upsampling and filtering network

According to this process, the considered LR frames are given as input to the filter generation network. The trained network generates the upsampling filters F as $\times H \times r^2$. The generated upsampling filter can be used for generating new pixels in filtered HR frame.

Finally, the HR pixels are generated by applying local filtering on the input X_t frame. Thus, the filtered output can be generated as:

$$\hat{Y}_t(yr+v, xr+u) = \sum_{j=-2}^{2}\sum_{i=-2}^{2} F_t^{x,y,v,u}(j+2, i+2) X_t(y+j, x+i) \quad (1)$$

Where x and y are the LR frame coordinates, v and u are the coordinates in each output block.

C. Feature Alignment Module

Let us consider that feature for each video frame are denoted as $\mathbb{F}_{t+i}, i \in [-N:+N]$. The convolution for K sampling location is computed where weight and offsets are defined as w_k and p_k for k locations. In this module, we consider 3×3 kernel with $K = 9$ and offset are given as $p_k \in \{(-1,1), (-1,0), \ldots, (1,1)\}$. The aligned feature $\mathbb{F}^a_{t+i}$ for each offset position can be expressed as:

$$\mathbb{F}^a_{t+i}(p_0) = \sum_{k=1}^{K} \mathbb{F}_{t+i}(p_0 + p_k + \Delta p_k).w_k \quad (1)$$

Where Δp is the learnable offset. In order to generate the l^{th} level pyramid features, we use stride convolutional filter which helps to down sample the features at $(l-1)^{th}$ pyramid by factor 2. Similarly, the offset and pyramid features can be predicted using upsampling the offset and aligned features from $(l+1)$ level.

$$\Delta P^l_{t+i} = f\left([\mathbb{F}_{t+i}, \mathbb{F}_t], \left(\Delta P^{l+1}_{t+i}\right)^{\uparrow 2}\right)$$
$$\mathbb{F}^a_{t+i}(p_0) = g\left(Conv\left(\mathbb{F}^l_{t+i}, \Delta P^l_{t+1}\right), \left((\mathbb{F}^a_{t+i})^{l+1}\right)^{\uparrow 2}\right) \quad (2)$$

D. Network Design

The proposed network architecture shares the weights to reduce the complexity. Moreover, we replace 2D convolution network with 3D layers to improve the learning process of spatio-temporal features. The complete network is constructed using batch normalization, ReLU, 1 scale order convolution, BN, ReLU and 3 scale convolution. In this process, each frame is processed through the proposed network architecture and obtained outcome of convolution is concatenated along with temporal axis. In order to generate the final HR frame, the filtered output and residual are added. The complete network architecture is depicted in figure 3.

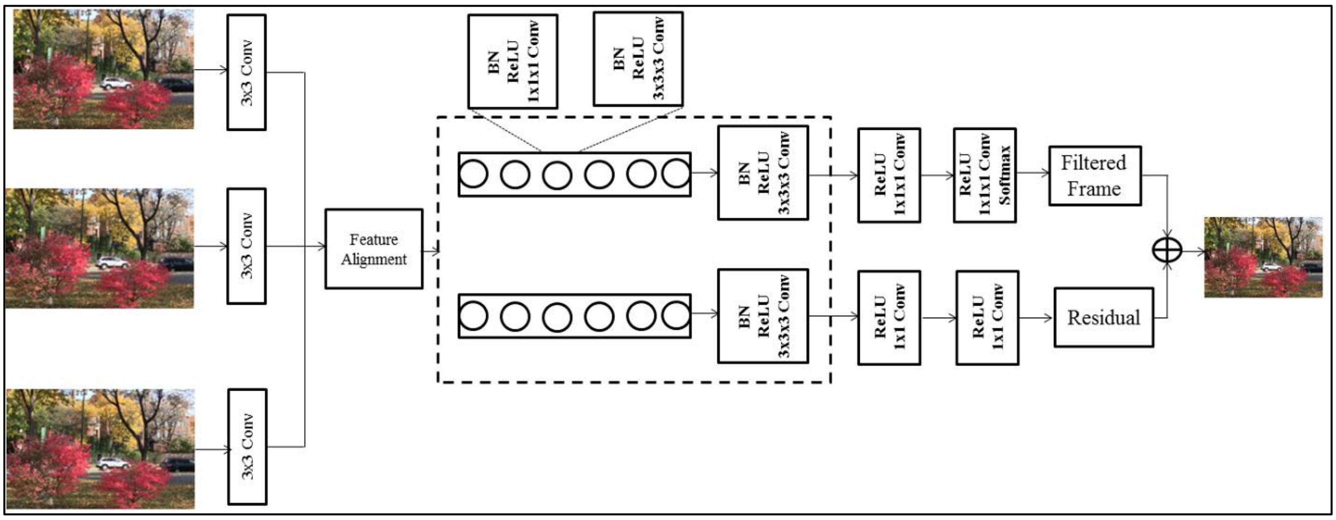


Fig.3. Proposed AFLSVR network architecture

IV. Results and Discussion

In this section, we present the brief discussion regarding datasets, training the database and comparative experimental analysis. The performance of proposed approach is compared with the existing techniques.

A. Dataset Description

In order to evaluate the performance of proposed approach, we consider open source video dataset such as REDS [26] and Vid4 [27]. REDS is a newly proposed high quality video data which has 240 training clips, 30 validation clips and 30 testing clips. Each video has 100 frames.

B. Comparative Analysis

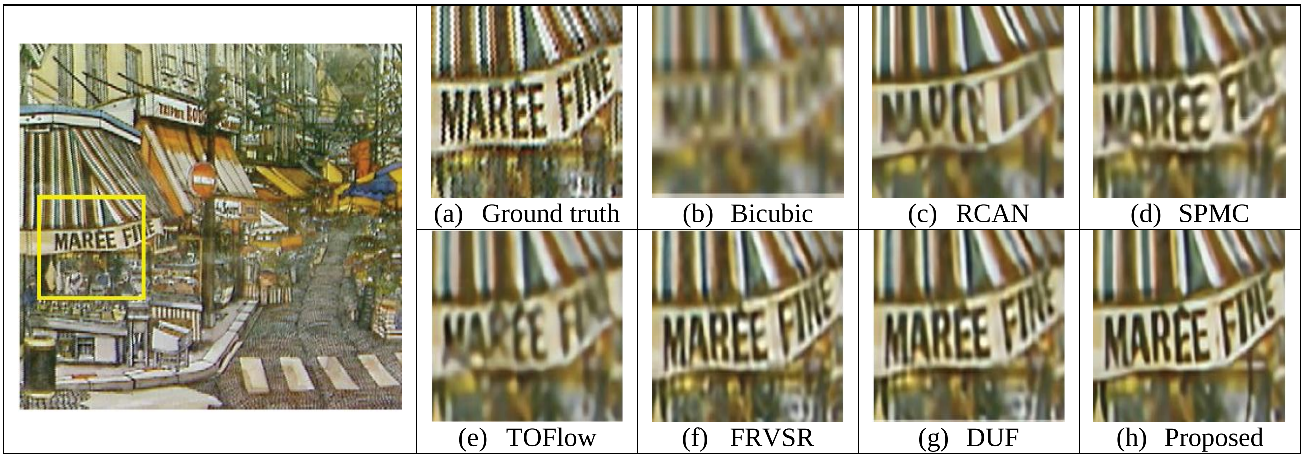
The performance of proposed AFLVSR is compared with the existing techniques such as Bicubic [28], RCAN [27], VESPCN [14], SPMC [17], TOFlow [30], FRVSR [21], DUF [31], and RBPN [22]. Below given table 2 shows quantitative comparative analysis for Vid4 dataset. The performance is measured in terms of PSNR and SSIM.

Table.2. PSNR and SSIM comparisons for Vid4 dataset

Clip Name	Bicubic (PSNR/SSIM)	RCAN (PSNR/SSIM)	VESPCN (PSNR/SSIM)	SPMC (PSNR/SSIM)	TOFlow (PSNR/SSIM)	FRVSR (PSNR/SSIM)	DUF (PSNR/SSIM)	RBPN (PSNR/SSIM)	Proposed
Calendar	20.39/0.572	22.33/0.7254	NA	22.16/0.746	22.47/0.731	NA	24.04/0.811	23.99/0.807	24.85/0.831
City	25.16/0.602	26.10/0.6960	NA	27.00/0.757	26.78/0.740	NA	28.27/0.831	27.73/0.803	28.20/0.8358

Foliage	23.47/0.566	24.74/0.6647	NA	25.43/0.720	25.27/0.709	NA	26.41/0.770	26.22/0.757	26.94/0.799
Walk	26.10/0.797	28.65/0.8719	NA	8.91/0.876	29.05/0.879	NA	30.60/0.914	30.70/0.909	31.32/0.952
Average	23.78/0.634	25.46/0.7395	25.35/0.7557	25.88/0.7752	25.89/0.765	26.69/0.822	27.33/0.831	27.12/0.818	28.05/0.912

According to the table 2, proposed approach achieves improved performance when compared with state-of-art techniques. For video sequence "Calendar", the proposed approach achieve PSNR as 24.85 which shows that the PSNR performance of proposed model is improved by 17.94%, 10.14%, 10.82%, 9.57%,3.25% and 3.46% when compared with Bicubic, RCAN, SPMC, TOFlow, FRVSR, DUF and RBPN techniques The qualitative comparative analysis for Vid4 dataset is depicted in below given figure 4.



(a) Ground truth (b) Bicubic (c) RCAN (d) SPMC
(e) TOFlow (f) FRVSR (g) DUF (h) Proposed

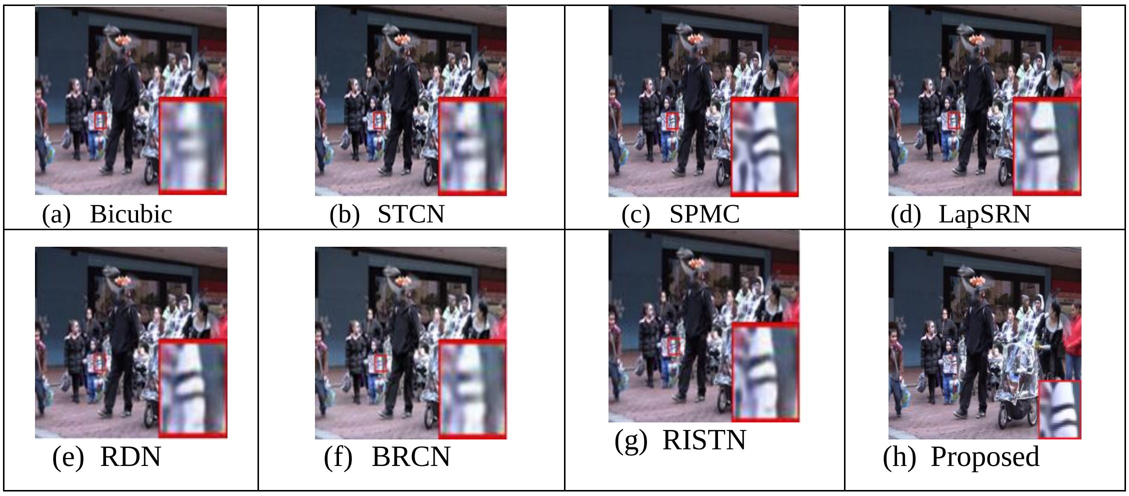
Fig.4. Qualitative comparisons of proposed approach with existing techniques

Similarly, we present a comparative analysis in terms of average PSNR and SSIM for Vid4 dataset. The details are mentioned in table 3. These values are obtained from [32].

Table.3. Average PSNR and SSIM performance comparisons

	Bicubic	VSRNet	STCN	VESPCN	SPMC	LapSRN	BRCN	RDN	RISTN	Proposed
PSNR	23.74	24.41	24.91	25.35	25.52	25.15	24.43	25.30	26.13	26.30
SSIM	0.633	0.707	0.734	0.757	0.760	0.771	0.712	0.712	0.792	0.833

The qualitative performance for Vid4 dataset is presented in figure 5 where we have considered "walk" sequence with scaling factor 4. The qualitative comparative analysis for "walk" sequence is depicted in figure 5.



(a) Bicubic (b) STCN (c) SPMC (d) LapSRN
(e) RDN (f) BRCN (g) RISTN (h) Proposed

Fig.5. qualitative comparisons of proposed approach with existing techniques

Similar to this experiment, we consider other frames of the given dataset and presented qualitative comparative analysis. This experiment is presented in figure 6 where we compare the "city", "calendar" and "foliage" sequences from Vid4 dataset. Below are the pictorial representation of results for different techniques and proposed scheme.

Frame	Bicubic	Liu et al. [16]	DUF [31]	WDVR [33]	Proposed

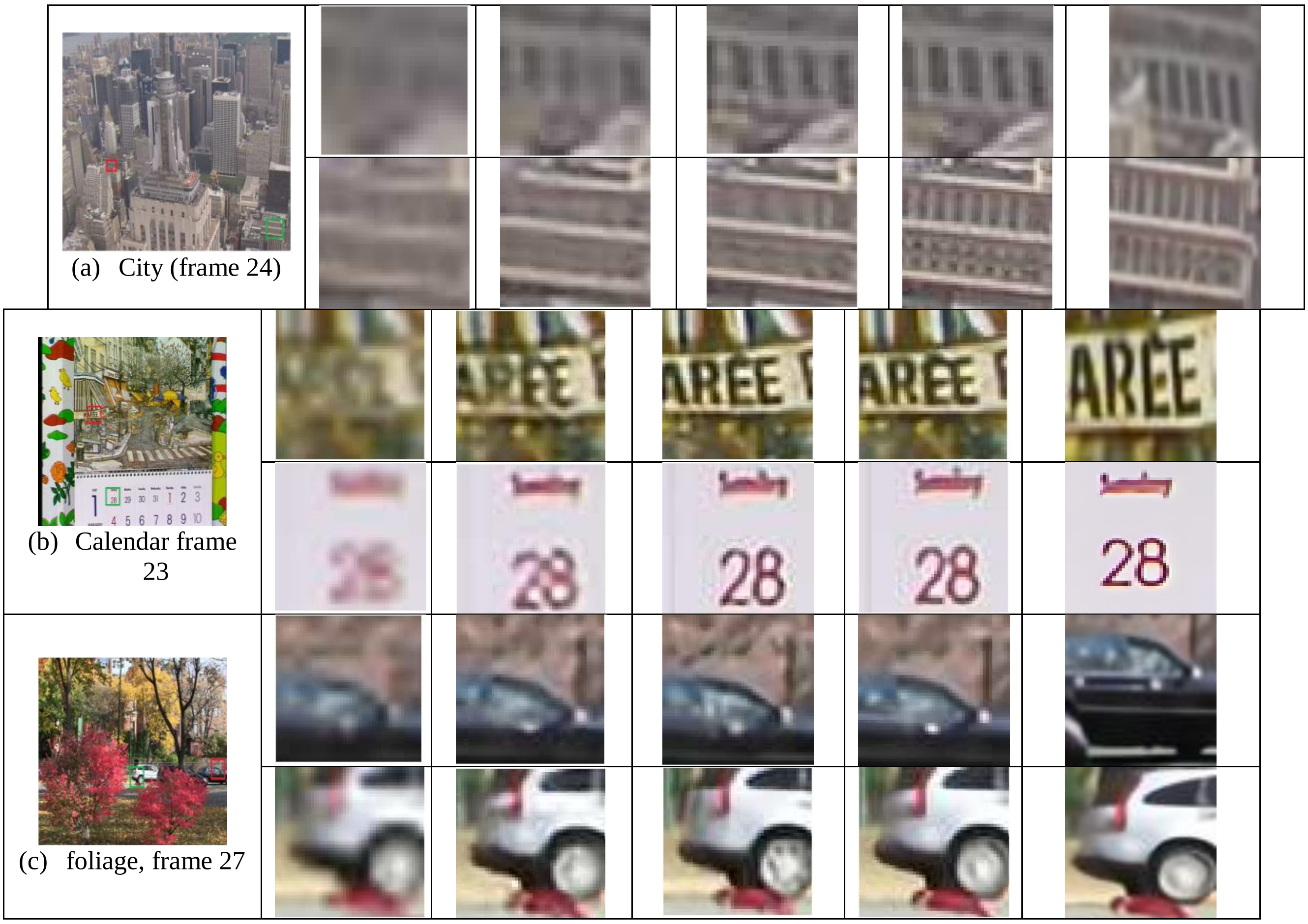


Fig.6. Qualitative results for Vid4 dataset.

Further, we measure the PSNR and SSIM performance for REDS4 dataset and compare the obtained performance with state-of-art techniques. The comparative analysis is presented table 4.

Table.4. PSNR and SSIM performance comparisons of various clips

Method	Clip_000 (PSNR/SSIM)	Clip_011 (PSNR/SSIM)	Clip_015 (PSNR/SSIM)	Clip_020 (PSNR/SSIM)	Average (PSNR/SSIM)
Bicubic	24.55/0.6489	26.06/0.7261	28.52/0.8034	25.41/0.7386	26.14/0.729
RCAN	26.17/0.7371	29.34/0.8255	31.85/0.8881	27.74/0.8293	28.78/0.8200
TOFlow	26.52/0.7540	27.80/0.7858	30.67/0.8609	26.92/0.7953	27.98/0.799
DUF	27.30/0.7937	28.38/0.8056	31.55/0.8846	27.30/0.8164	28.63/0.8251
Proposed	28.45/0.846	32.17/0.8864	34.70/0.9380	31.18/0.9021	31.62/0.8931

According to this experiment, proposed approach achieves average PSNR as 31.62 and SSIM as 0.8931 which shows a significant improvement compared to existing techniques.

V. CONCLUSION

In this article, we present a novel architecture to improve the video resolution [29]. Traditionally, single image based super-resolution techniques have been introduced but these techniques fail to achieve the desired performance for video frames because these techniques are not able to exploit complete spatial and temporal information between inter-frames. Moreover, inappropriate feature alignment may lead towards the blurry frame generation. In this work, we have introduced deep leaning based architecture which uses feature alignment model, convolutional, batch normalization modules to construct the image filtering and image residual architecture. Finally, the HR frame is constructed by combining filtered and residual frames. In future, this work can be extended towards video Super resolution to address the blurriness issues in low-resolution frames.

REFERENCES

1. Zhang, K., Zuo, W., & Zhang, L. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. IEEE Transactions on Image Processing, 27(9), 4608-4622.
2. Yang, X., Xu, K., Song, Y., Zhang, Q., Wei, X., & Lau, R. W. (2018). Image correction via deep reciprocating HDR transformation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1798-1807).
3. Sahu, S., Lenka, M. K., & Sa, P. K. (2019). Blind Deblurring using Deep Learning: A Survey. arXiv preprint arXiv:1907.10128.
4. Wang, Z., Chen, J., & Hoi, S. C. (2019). Deep learning for image super-resolution: A survey. arXiv preprint arXiv:1902.06068.
5. Wang, Z., Yi, P., Jiang, K., Jiang, J., Han, Z., Lu, T., & Ma, J. (2018). Multi-memory convolutional neural network for video super-resolution. IEEE Transactions on Image Processing, 28(5), 2530-2544.
6. Babacan, S. D., Molina, R., & Katsaggelos, A. K. (2010). Variational Bayesian super resolution. IEEE Transactions on Image Processing, 20(4), 984-999.
7. Zhu, X., Wang, X., Wang, J., Jin, P., Liu, L., & Mei, D.

(2017). Image Super-Resolution Based on Sparse Representation via Direction and Edge Dictionaries. Mathematical Problems in Engineering, 2017.
8. Qiqin Dai, Seunghwan Yoo, Armin Kappeler, and Aggelos K Katsaggelos. Dictionary-based multiple frame video superresolution. In ICIP, 2015.
9. Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In CVPR, 2017
10. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2472-2481).
11. Lin, G., Wu, Q., Qiu, L., & Huang, X. (2018). Image super-resolution using a dilated convolutional neural network. Neurocomputing, 275, 1219-1230.
12. Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295-307.
13. Bare, B., Yan, B., Ma, C., & Li, K. (2019). Real-time video super-resolution via motion convolution kernel estimation. Neurocomputing.
14. J. Caballero , C. Ledig , A. Aitken , A. Acosta , J. Totz , Z. Wang , W. Shi , Real-time video super-resolution with spatio-temporal networks and motion compensation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2848–2857 .
15. D. Li , Z. Wang , Video super-resolution via motion compensation and deep residual learning, IEEE Trans. Comput. Imaging PP (99) (2017) 1 .
16. D. Liu , Z. Wang , Y. Fan , X. Liu , Z. Wang , S. Chang , T. Huang , Robust video su20per-resolution with learned temporal dynamics, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2526–2534.
17. X. Tao , H. Gao , R. Liao , J. Wang , J. Jia , Detail-revealing deep video super-resolution, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 4 482–4 490 .
18. Kappeler, A., Yoo, S., Dai, Q., & Katsaggelos, A. K. (2016). Video super-resolution with convolutional neural networks. IEEE Transactions on Computational Imaging, 2(2), 109-122.
19. Liu, D., Wang, Z., Fan, Y., Liu, X., Wang, Z., Chang, S., & Huang, T. S. (2018). Learning temporal dynamics for video super-resolution: A deep learning approach. IEEE Transactions on Image Processing, 27(7), 3432-3445.
20. Wang, X., Chan, K. C., Yu, K., Dong, C., & Change Loy, C. (2019). EDVR: Video restoration with enhanced deformable convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.
21. Sajjadi, M. S., Vemulapalli, R., & Brown, M. (2018). Frame-recurrent video super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 6626-6634).
22. Haris, M., Shakhnarovich, G., & Ukita, N. (2019). Recurrent Back-Projection Network for Video Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3897-3906).
23. Li, S., He, F., Du, B., Zhang, L., Xu, Y., & Tao, D. (2019). Fast Spatio-Temporal Residual Network for Video Super-Resolution. arXiv preprint arXiv:1904.02870.
24. Yi, P., Wang, Z., Jiang, K., Shao, Z., & Ma, J. (2019). Multi-Temporal Ultra Dense Memory Network for Video Super-Resolution. IEEE Transactions on Circuits and Systems for Video Technology.
25. Du, X., Zhou, Y., Chen, Y., Zhang, Y., Yang, J., & Jin, D. (2019, July). Dense-Connected Residual Network for Video Super-Resolution. In 2019 IEEE International Conference on Multimedia and Expo (ICME) (pp. 592-597). IEEE.
26. Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenges on video deblurring and superresolution: Dataset and study. In CVPRW, June 2019. 1, 5
27. Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In ECCV, 2018.
28. Renjie Liao, Xin Tao, Ruiyu Li, Ziyang Ma, and Jiaya Jia. Video super-resolution via deep draft-ensemble learning. In CVPR, 2015
29. Ce Liu and Deqing Sun. On bayesian adaptive video super resolution. TPAMI, 36(2):346–360, 2014
30. Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In CVPR, 2018
31. Younghyun Jo, Seoung Wug Oh, Jaeyeon Kang, and Seon Joo Kim. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. In CVPR, 2018
32. Zhu, X., Li, Z., Zhang, X. Y., Li, C., Liu, Y., & Xue, Z. (2019, July). Residual Invertible Spatio-Temporal Network for Video Super-Resolution. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 5981-5988).
33. Fan, Y., Yu, J., Liu, D., & Huang, T. S. (2019). An Empirical Investigation of Efficient Spatio-Temporal Modeling in Video Restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.

AUTHORS PROFILE

Padma Reddy A.M is a Research Scholar in "School of Computing & Information Technology", Reva University, Bangalore. He has over 2 decades of teaching experience. Currently he is working as a Professor and Dean Admissions and Student Welfare, at SVIT, Bangalore. His research interests are video processing, N/w and language translators

Dr. Udaya Rani V received her PhD from the Mother Teresa University, Tamil Nadu, India in 2014. She is working at REVA University as Sr. Associate Professor since 2011. She has approx. 17yrs of experience and published above 20 papers in National and International Journals. Her research area includes Data Mining and Warehousing, Machine Learning and Big Data.