

PAPER • OPEN ACCESS

## An Overview of Video Super-Resolution Algorithms

To cite this article: Chenming Liu *et al* 2021 *J. Phys.: Conf. Ser.* **2025** 012051

View the [article online](#) for updates and enhancements.

### You may also like

- [An Improved Training Sample Set Construction Method](#)  
Li Shen and Zhai Jiaojiao
- [Video Super-Resolution using Multi-Frames Fusion and Perceptual Loss](#)  
Xiaonan Zhu, Yucong Zhao, Ping Wang et al.
- [The Precision Exploration of Image Super-Resolution Applied into Motion Measurement](#)  
Ye Wei, Wenhui Hou, Wei Feng et al.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology

## 241st ECS Meeting

May 29 – June 2, 2022 Vancouver • BC • Canada  
Abstract submission deadline: Dec 3, 2021

Connect. Engage. Champion. Empower. Accelerate.  
**We move science forward**



**Submit your abstract**



# An Overview of Video Super-Resolution Algorithms

Chenming Liu<sup>1,2</sup>, Ruipeng Gang<sup>1</sup>, Jinjing Li<sup>2</sup>, Ji Fang<sup>1</sup> and Haodong Yu<sup>3</sup>

<sup>1</sup> Academy of Broadcasting Science, NRTA, Beijing 100866, China

<sup>2</sup> State Key Laboratory of Media Convergence and Communication, Beijing 100024, China

<sup>3</sup> Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany  
Email: liuchenming@abs.ac.cn; gangruipeng@abs.ac.cn; 1792418414@qq.com

**Abstract.** We investigate some excellent algorithms in the field of video space super-resolution based on artificial intelligence, structurally analyze the network structure of the algorithm and the commonly used loss functions. We also analyze the characteristics of algorithms in the new field of video space-time super-resolution. This work helps researchers to deeply understand the video super-resolution technology based on artificial intelligence.

**Keywords.** Video super-resolution; deep learning; convolutional neural network; recurrent neural network.

## 1. Introduction

Video super-resolution technology is a basic technology in the field of video enhancement and repair. Before the rise of deep learning, researchers used traditional image processing methods to achieve video resolution, like up-conversion. In the 1960s, Harris and Goodman [1-5] proposed to recover the image information beyond the limit frequency of the optical system modulation transfer function (MTF) by spectral extrapolation, which was the origin of image super-resolution algorithm and video super-resolution algorithm. At present, the commonly used restoration methods include bilinear interpolation, local adaptive amplification interpolation, cubic spline interpolation, etc. To process video in the way of image processing, video should be first parsed into video frames, and then each video frame should be handled by image super-resolution algorithm. This simple image interpolation method ignores the timing dimension of the video, and fails to take into account the correlation between the frames before and after the video and the blurring caused by the video transition and fast movement [6]. Although the restoration speed is very fast, the final reconstructed high-resolution video will produce incoherent, fuzzy, and other degraded effects, leading to the unideal subjective effect.

Currently, video super-resolution algorithms based on Artificial Intelligence (AI) have become a trending technology in the field of AI video enhancement/repair. With the successful application of deep learning in spatial super-resolution of video, the space-time super-resolution technology derived from video super-resolution technology realizes the up-conversion from low definition to ultra-high resolution, and from low frame rate to high frame rate. At present, the spatial super-resolution algorithm and spatial-temporal super-resolution algorithm of AI video are generally based on Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Recurrent Neural Network (RNN) [7-8]. Compared with traditional algorithms, video super-resolution technology based on AI takes the temporal dimension into account, and it pays more attention to the connection between the video reference frame and adjacent frames in the processing process. Especially when dealing with occlusion, large-scale motion and severe blur, it is necessary to align multiple frames and establish



accurate correspondence, as well as effectively fuse the aligned features for reconstruction. Therefore, the AI algorithm can do a better job of restoring the video.

## 2. Video Spatial Super-Resolution Algorithm Based on Deep Learning

### 2.1. Input and Output of Network

The input of network is generally divided into two categories: reference frame and adjacent frame. Multiple consecutive video frames are used as input, such as EDVR [9], RBPN [10], FISR [11] and MuCAN [12]. These models generally input multiple video frames into their respective preprocessing modules, and then get the features containing a lot of information. Base on that, the model can get much more ideal experimental results.

Theoretically, more frames input at one time, more information contained, and better results obtained after recovery. For example, TGA [13] will take 7 consecutive video frames as input at one time and output 1 video frame at last. This model has achieved good results on the test dataset. Some network models choose 2 video frames as input, such as Zooming Slow-Mo [8] and STARNet [14], which are based on RNN network and can output 2 corresponding high-resolution frames at the same time. In addition, STARNet uses RBPN as a pre-training model. As the result, for videos with many motion scenes, STARNet has a good generalization ability, the experimental results are significant.

### 2.2. Network Structure

The structure of spatial super-resolution network can be divided into frame alignment, feature extraction and fusion, attention mechanism, and reconstruction. Frame alignment is the key to determine the effect of model experiment. Feature extraction fusion can extract the main information of data. The attention mechanism can further distinguish the importance of different features in the main messages. In general, the existing image super-resolution algorithm structure is used to perform reconstruction work such as up-transform.

### 2.3. Frame Alignment Module

Before the continuous video frames are input into the network model, it is necessary to align the adjacent frames to the reference frame by means of frame alignment, that is, adding the time information of adjacent frames to the reference frame, then the input frames are fused with features.

There are many methods of frame alignment, which can be categorized as explicit alignment and implicit alignment. Explicit alignment is pixel level alignment on the original image, using optical flow method through motion estimation and compensation generally (VESPCN [15], TOFLOW [16], RBPN, SOFVSR [17]). The optical flow method calculates the optical flow by pre-training an optical flow model and aligning adjacent frames to reference frames so as to realize information reuse. The optical flow method will usually first operate at low resolution, which can reduce a certain amount of computation. Of course, there are also algorithms for alignment at high resolution conditions, such as TecoGAN [18], FRVSR, etc. These algorithms calculate the optical flow between previous frame and current frame, and apply the up-sampling directly to the generated results of the previous frame, so as to realize the reuse of high-resolution images.

Although algorithms using explicit alignment module is widely used in video super-resolution, the algorithms also have some problems. (1) The quality of video frame alignment depends heavily on the pre-trained optical flow model, which increases the workload and makes the algorithm are not strong robustness. (2) Motion occlusion and other problems often exist in moving videos, which can not be well solved by optical flow method. Problems such as blur and flicker are likely to occur, which can easily affect the performance of the model.

Implicit alignment is usually performed at the feature level, the deformable convolution is representative of implicit alignment algorithms. TDAN [19] first applied deformable convolution to video super-resolution, and DNLN and EDVR also applied it with good results. Since deformable convolution can adaptively select regions of interest to itself, this makes it possible for variable

convolution to extract regions that are useful for video frame reconstruction to a certain extent. However, before performing deformable convolution, the offsets of two sets of features need to be calculated using normal convolution and input to the deformable convolution unit together with the adjacent frames, and the reliance on the calculation of the offsets makes the network less stable during the training process.

#### 2.4. Feature Extraction and Fusion

Feature extraction and fusion is an effective way to extract key information, reduce repetitive information in consecutive video frames, and speed up the computational efficiency of the model.

CNN algorithms have the ability of representation learning and classify the input data in terms of translation invariance [20]. The VESPCN, DRVSR [21] algorithms will use CNN for feature extraction and fusion. For example, VESPCN will input motion-compensated video frames into a series of CNN layers and then get the features of the video frames to be reconstructed; DRVSR is designed a feature fusion module, which first down-sampled the image using a convolution kernel with a step size of two, and then used deconvolution to obtain a HR residual image, which can help the model to recovery to super-resolution; The residual module is also a widely used model, in MMCNN [22], where a feature extraction method is designed based on residual dense networks. With the residual module, the image information is not easily lost in the transmission of the network and the model can obtain rich features.

#### 2.5. Reconfiguration Modules

To overcome the shortcomings of traditional methods and to achieve end-to-end behavior of the algorithm, two common reconstruction methods are presented here: Transposed Convolution [23] and Sub-pixel Convolution [24].

The difference between the deconvolution and the convolution in CNN is that the convolution in CNN is based on the input feature image information to get the corresponding features, while the deconvolution can output the final image based on a set of features learned by the model. Taking a 2x hyper-segmented and 3x3 feature map as an example, the input feature map will first be expanded by a factor of 2.

After computing the enlarged feature map by using convolution kernel of size 3x3 and set padding to 1, an image with a magnification of 2 was obtained. Because deconvolution achieves image enlargement in an end-to-end manner, it is also widely used in various reconstruction modules. However, deconvolution can also easily cause uneven overlap on different axes, which ultimately affects the reconstructive effect of the model.

Sub-pixel convolution is another end-to-end reconstruction method, where multiple channels of feature maps are generated by convolution, and then the shape of this set of feature maps has changed.

Suppose we have a 3x3 feature map, in order to obtain the corresponding 2x image, we first compute the feature map of 4 channels using convolution, and then fuse this set of feature maps into a single channel image, which gives us the result after 2x super-resolution. Due to the end-to-end reconstruction approach, sub-pixel convolution is also used by various video super-resolution models. Compared to the inverse convolution layer, sub-pixel convolution has a larger receptive field, providing more contextual information, and helping to generate more realistic details. However, as the receptive fields are not evenly distributed, block regions actually share the same receptive field, which may lead to some artefacts near the boundaries of different blocks. On the other hand, independent prediction of adjacent pixels in a block region may lead to an unsmooth output. Therefore, Gao and his team [25] proposed PixelTCL, which replaces interdependent sequential prediction to produce smoother and more consistent results.

#### 2.6. Attention Mechanisms

Attention mechanisms are widely used in video super-resolution algorithms. EDVR provides a time-space attention fusion module, which can calculate the similarity between video frames in a new space. And for video frames with higher similarity, this model can give more attention by this module; DNLN designs a non-local attention mechanism module based on deformable convolution, which can

significantly improve the effect after video reconstruction; TGA [13] designed a time-domain grouped attention module that reduces the redundant information of the input, and can also bring an improvement in the effect.

### 2.7. Other Modules

In addition to the common modules, scholars have also proposed many other novel superpartition modules, which can effectively improve the model effect.

RBNP, based on the idea of RNN, designs a mapping module combining an encoder and a decoder. The encoder mainly consists of image super-resolution related model structure, and the decoder mainly consists of the residual network. TecGAN [18] designed a novel network on the basis of GAN, which is able to estimate the optical flow between reference frames and adjacent frames. It used bicubic interpolation to amplify the corresponding HR optical flow map, and combined the HR optical flow map with the GT of the previous frame to obtain GT', and finally fed both GT' and GT into the discrimination network for discrimination, so as to alleviate smoothing over and non-smoothness of the output results.

### 2.8. Loss Function

The types of loss functions can be classified into pixel-level loss, structural loss, content loss, and adversarial loss. For video super-resolution, pixel-level loss and structural loss are mainly used as evaluation metrics. And to further enhance the subjective visual effect, a weighted loss of "pixel-level + structural + content" is generally used.

### 2.9. Pixel Loss

(1) L1 & L2 loss

$$L_1 = \frac{1}{N} \sum_{p \in P} |y'(p) - y(p)| \quad (1)$$

$$L_2 = \frac{1}{N} \sum_{p \in P} (y'(p) - y(p))^2 \quad (2)$$

Here  $N$  represents the number of pixels, and  $p$  represents all pixel positions in the graph.  $y'$  is the generated image and  $y$  is the ground truth.

Practically, both L1 and L2 have their own drawbacks. For L1 is that the gradient is not smooth, which tends to make the network training unstable. L2 will increase the punishment intensity for large errors, while it is opposite for small errors. Therefore, gradient explosion is very easy to occur at the beginning of network training. In order to alleviate the above problems, the loss functions of L1 and L2 are alternately used to optimize the training of the network in super-resolution models. The loss function of ClipL1 is basically the same as the loss of L1, but the gradient trimming method can control the magnitude of the gradient and make the training process more stable during the back propagation.

(2) Charbonnier Loss

$$\text{Charbonnier}(x, y) = \sqrt{(x - y)^2 + \varepsilon^2} \quad (\varepsilon \text{ is usually equal to } 1e-6) \quad (3)$$

Charbonnier loss avoid excessive gradient caused by excessive processing outliers, so that the generated effect can be clearly defined and cleared. The loss was first applied in the LAPSRN.

### 2.10. Structural Loss

(1)SSIM Loss

SSIM measures three factors: luminance, contrast, and structure.

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$$

$$c(x, y) = \frac{2\delta_x\delta_y + c_2}{\delta_x^2 + \delta_y^2 + c_2}$$

$$s(x, y) = \frac{\delta_{xy} + c_3}{\delta_x\delta_y + c_3}$$

Here  $\mu$  is the mean,  $\sigma$  is the variance, and  $\delta_{xy}$  is the covariance of  $x$ ,  $y$ .  $c_1 = (k_1L)^2$  and  $c_2 = (k_2L)^2$  are constants to avoid dividing by zero, where  $L$  is the range of pixel value,  $k_1$  is 0.01 and  $k_2$  is 0.03.

$$\text{SSIM}(x, y) = [l(x, y)^\alpha \cdot (x, y)^\beta \cdot s(x, y)^\gamma] (\alpha, \beta, \gamma \text{ are usually equal to } 1) \quad (4)$$

SSIM is one of the most commonly used loss functions in video super-resolution algorithms, taking into account the fact that human eyes are sensitive to structural information, whereas insensitive to distortion in high luminance areas and more contrast areas of “texture”.

$$\begin{aligned} l(x, y) &= \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} c(x, y) = \frac{2\delta_x\delta_y + c_2}{\delta_x^2 + \delta_y^2 + c_2} s(x, y) = \frac{\delta_{xy} + c_3}{\delta_x\delta_y + c_3} \text{SSIM}(x, y) \\ &= [l(x, y)^\alpha \cdot (x, y)^\beta \cdot s(x, y)^\gamma] \end{aligned}$$

### 2.11. Content Loss

Content loss function estimates the perceived quality of images. Generally, pre-trained image classification network is used to measure the differences in image content, which can be expressed as:

$$L_{content}(I', I; \Phi, l) = \frac{1}{h_l w_l c_l} \sqrt{\sum_{i,j,k} (\Phi_{i,j,k}^{(l)}(I') - \Phi_{i,j,k}^{(l)}(I))^2} \quad (5)$$

$h_l$ ,  $w_l$ , and  $c_l$  represent the height, width, and channel number of the first layer respectively;  $\Phi$  represents the pre-trained model (VGG and RESNET are commonly used);  $I'$  and  $I$  represent the generated image and the real image. The essence of content loss is to transfer the hierarchical features of image in the classification network to the super-resolution reconstruction network.

### 2.12. Adversarial Loss.

Among the super-resolution algorithms related to GAN, the loss function of WGAN is used most.

$$L_{gan\_ce\_g} = -\log(\hat{I}) \quad (6)$$

$$L_{gan\_ce\_d} = -\log(I_s) - \log(1 - D(\hat{I})) \quad (7)$$

where  $L_{gan\_ce\_g}$  and  $L_{gan\_ce\_d}$  denote the adversarial loss of the generator and the discriminator, and  $I_s$  represents images randomly sampled from the ground truths. The reconstruction network based on GAN will reduce the fidelity to some extent, as the result, the loss of strong monitoring algorithms such as PSNR and SSIM will not be applied to it.

## 3. Video Spatial-Temporal Super-Resolution Algorithm Based on Deep Learning

Video spatio-temporal super-resolution algorithm is algorithm that pays more attention to the combination of spatial and temporal information of video, which can improve the resolution and frame rate of video at the same time. Compared with the video spatial super-resolution algorithm, the video spatial super-resolution algorithm is disparate in model structure, but similar in structure, loss function, and etc. At present, the video spatio-temporal super-resolution model is mainly composed of basic structures such as 3D convolution and Recurrent Convolutional Neural Network (RCNN).

### 3.1. Algorithm Based on 3D Convolution

Compared with 2D convolution, 3D convolution increases the time sequence dimension and can independently extract the spatio-temporal correlation between consecutive frames, so it is very suitable for processing video data. DUF [26], FSTRN [27], and 3DSRnet [28] all use 3D convolution based methods to integrate video time and space information.

DUF adopts the dynamic change network structure, which can generate different convolution kernels according to different inputs, so as to obtain the adaptive feature map. DUF uses 3D convolution to independently extract the time information between consecutive frames, making the output of the model more stable. FSTRN uses 3D convolution for feature extraction and feature fusion, where a  $k \times k \times k$

convolution kernel is divided into two cascaded convolution kernels with sizes of  $1 \times k \times k$  and  $k \times 1 \times 1$  respectively, which can effectively reduce the computational load of the model. 3DSRnet extracts features from continuous video frames through 3D convolution, and provides a scene transformation method, which replaces frames with the nearest adjacent frames of the same scene when the scene changes. This method can overcome the performance degradation caused by scene changes to a certain extent. However, compared with 2D convolution methods, most 3D convolution methods have higher computational complexity, which limits their application for real-time video super-resolution tasks.

### 3.2. RCNN-Based Algorithm

RCNN is a combination of RNN and CNN, taking advantage of RNN's ability to handle serialization. Many models, such as STCN [29] and RISTN [30], adopt the idea of RCNN to solve the video super-resolution problem.

STCN is an algorithm without motion estimation and motion compensation. STCN consists of three parts: space processing module, time processing module, and reconstruction module. Space module is for feature extraction of continuous low resolution video frames; Time processing module is a bidirectional convolution neural network, which is mainly used to extract features in the video time domain; RISTN which is used to reconstruct consists of three modules: reversible residual block, LSTM (Long Short-Term Memory) combined with dense residual network, RDC-LSTM, and sparse feature fusion module. Among them, RDC-LSTM can be used to extract the temporal and spatial information of the video. After extracting the temporal and spatial information, the sparse fusion strategy is adopted to selectively fuse the features. However, the traditional RCNN based method is difficult to train and sometimes has the problem of gradient disappearance. When the length of input sequence is too large, the long-term dependence of the input may not be captured, and good performance cannot be obtained.

## 4. Conclusion and Future Directions

So far, video super-resolution technology based on deep learning is still limited to hardware equipment conditions, and single image super-resolution. Although the optical flow method has been widely tested, and highly innovative models such as EDVR have been constantly introduced, the development of deep learning-based video super-resolution technology has not yet met the expectations of developers. At the same time, there are some problems that need to be overcome, such as the processing of the front and back frames during the special video session, the processing of subtitles, and the resolution of the super-resolution results when replayed, etc. There is still a long way to go for deep learning based video super-resolution technology.

## Acknowledgement

This work was supported in part by Beijing Municipal Science & Technology Commission (No. Z191100003419003) and Academy of Broadcasting Science, NTRA (ZZLX-2020-001).

## References

- [1] Harris J L 1964 Diffraction and resolving power *Journal of the Optical Society of America* **54** (7) 931-933.
- [2] Suresh S, Babu R V and Kim H J 2008 No-reference image quality assessment using modified extreme learning machine classifier *Applied Soft Computing Journal* **9** (2) 541-552.
- [3] Criminisi A, Perez P and Toyama K 2004 Region filling and object removal by exemplar-based image inpainting *IEEE Transactions on Image Processing* **13** (9) 1200-1212.
- [4] Freeman W T, Pasztor E C and Carmichael O T 2000 Learning low-level vision *International Journal of Computer Vision* **40** (1) 25-47.
- [5] Freeman W T, Jones T R and Pasztor E C 2002 Example-based super-resolution *Computer Graphics & Applications IEEE* **22** (2) 56-65.
- [6] Wang Z, Chen J and Hoi S 2019 Deep learning for image super-resolution: A survey *IEEE Transactions on Pattern Analysis and Machine Intelligence* **40** (99) 1-1.

- [7] Chu M, Xie Y and Mayer J 2020 Learning Temporal Coherence via Self-Supervision for GAN-based Video Generation *ACM Transactions on Graphics (TOG)* **39** (75) 1 -13.
- [8] Xiang X, Tian Y, Zhang Y, Fu Y, Allebach J P and Xu C 2020 Zooming Slow-Mo: Fast and Accurate One-Stage Space-Time Video Super-Resolution *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 3367-3376.
- [9] Wang X, Chan K C K, Yu K, Dong C and Loy C C 2019 EDVR: Video restoration with enhanced deformable convolutional networks *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* pp 1954-1963.
- [10] Haris M, Shakhnarovich G and Ukita N 2019 Recurrent back-projection network for video superresolution *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp 3892-3901.
- [11] Zhang K and Gool L V 2020 Deep unfolding network for image super-resolution *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp 3214-3223.
- [12] Isobe T, Jia X and Gu S 2020 Video super-resolution with recurrent structure-detail network *European Conference on Computer Vision* pp 645-660.
- [13] Isobe T, Li S and Jia X 2020 Video super-resolution with temporal group attention *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp 8005-8014.
- [14] Haris M, Shakhnarovich G and Ukita N 2020 Space-time-aware multi-resolution video enhancement *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp 2859-2868.
- [15] Caballero J, Ledig C and Aitken A 2017 Real-time video super-resolution with spatio-temporal networks and motion compensation *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* pp 2848-2857.
- [16] Xue T, Chen B, Wu J, Wei D and Freeman W 2018 Video enhancement with task-oriented flow *International Journal of Computer Vision* 1-20.
- [17] Wang L, Guo Y, Lin Z, Deng X, An W 2018 Learning for Video Super-Resolution through HR Optical Flow Estimation *Asian Conference on Computer Vision* 514-529.
- [18] Sajjadi M S, Vemulapalli R and Brown M 2018 Frame-recurrent video super-resolution *IEEE/CVF Conference on Computer Vision and Pattern Recognition* pp 6626-6634.
- [19] Tian Y, Zhang Y and Fu Y 2020 TDAN: Temporally-deformable alignment network for video super-resolution *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp 3357-3366.
- [20] Zhang W 1988 Shift-invariant pattern recognition neural network and its optical architecture *Proceedings of Annual Conference of the Japan Society of Applied Physics*.
- [21] Tao X, Gao H, Liao R, Wang J and Jia J 2017 Detail revealing deep video super-resolution *IEEE Computer Society* pp 4482-4490.
- [22] Wang Z, Yi P, Jiang K, Jiang J, Han Z, Lu T and Ma J 2019 Multi-memory convolutional neural network for video super-resolution *IEEE Trans. Image Process* **28** (5) 2530-2544.
- [23] Zeiler M D, Krishnan D, Taylor G W and Fergus R 2010 Deconvolutional networks *Computer Vision & Pattern Recognition* 2528-2535.
- [24] Shi W, Caballero J, Huszár F 2016 Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 1874-1883.
- [25] Gao H, Yuan H and Wang Z 2019 Pixel transposed convolutional networks *IEEE Transactions on Pattern Analysis and Machine Intelligence* **42** (5) 1218-1227.
- [26] Jo Y, Oh S, Kang J and Kim S 2018 Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation *IEEE/CVF Conference on Computer Vision and Pattern Recognition* pp 3224-3232.
- [27] Li S, He F, Du B, Zhang L, Xu Y and Tao D 2019 Fast spatio-temporal residual network for video super-Resolution *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp 10514-10523.



- [28] Kim S Y, Lim J, Na T and Kim M 2019 Video super resolution based on 3D-CNNs with consideration of scene change *IEEE Int. Conf. Image Process* pp 2831-2835.
- [29] Guo J and Chao H 2017 Building an end-to-end spatial-temporal convolutional network for video super-resolution *AAAI* 4053-4060.
- [30] Mudenagudi U, Banerjee S and Kalra P 2009 On improving space-time super resolution using a small set of video inputs *Conference on Computer Vision, Graphics & Image Processing* pp 320-327.