International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

CCTV Surveillance Camera's Image Resolution Enhancement using SRGAN

[1]Prof. M. Seshaiah, [2]Abhijith R Nair, [3]Ektha Mallya, [4]Shiv Dev

[1]Assistant professor in Dept. of CS&E, VTU, Belgaum, India,
[2,3,4]Undergraduate student in Dept. of CS&E VTU, Belgaum, India

---

Abstract - *Computer Vision, Image enhancement, Artificial Intelligence are some of the fields that are being used for the improvement in the resolution of the images obtained from a surveillance camera. The surveillance camera is installed in almost every corner of the city and hence it can either be present internally or externally. There are many factors such as weather, light conditions and poor quality of the surveillance device, that affect the quality of the images or video being recorded. Hence the images recorded in such conditions require special attention and hence need to be enhanced for better results. Therefore in this paper, we have implemented the enhancement method for images that have a poor scale low resolution. Implemented method uses a machine learning algorithm, Super Resolution Generative Adversarial Networks (SRGAN) for achieving the goal of enhancement of images obtained from surveillance cameras. Super resolution of images allows us to obtain images with better resolution and less noise and hence provides the users with better experience of using the surveillance system.*

Key Words: **SRGAN, Image Enhancement, surveillance system, noise reduction, Super Resolution, Upsampling, Batch Normalisation, Peak Signal to Noise Ratio(PSNR)**

1. INTRODUCTION

The number of Closed-Circuit Television (CCTV) exponentially increases every year for various reasons. The reason could be due to the increasing crime rate, everything is put in a record which helps the crime department find the culprit based on proof and recognition of the event or to regulate traffic violation[1]. Hence the need for a surveillance system is always essential.

Our proposed survey work is focused on CCTV Image Enhancement and the use of SRGAN for the same. According to the survey analysis, the CCTV systems are the most widely used technology for monitoring various activities. The main limitations in the images obtained from the CCTV cameras are poor quality images. This could be due to many reasons such as there is high noise in the image and the information being carried is a degraded version of the original. Environmental issues such as fog, rain, snow distort the images and pose a threat to the quality of images [2].The images can also be compressed by the system at a low resolution and hence can be degraded due to noises, blurs or bad illumination[3].

These challenges have been overcome by using image enhancement techniques and algorithms such as using sub-images homomorphic filtering techniques where an image is divided horizontally and vertically and enhanced[4]; noise adaption super resolution which reserves some values like edges and when the noise in the image increases, the value preserved is utilized[5]; Convolutional Neural Network helps us use deep learning architecture which simplifies the model to high extent [6]; Deep Convolutional Generative Adversarial Networks allows us to simplify the architecture and helps us build a model which performs various image enhancement techniques such as image-super resolution, image-denoising and deconvolution that provides with optimum results for enhancing the images obtained from a CCTV system[7]; Feature Extraction which is a computer vision technique and is efficiently used for image filtering purposes[8] and hence is usually used as a preprocessing element.

For solving the limitations of the images procured from the surveillance system, a few methods have proved to give a faster and efficient result. Images before being fed into the model are preprocessed , that is the images are normalized to a particular size and are also converted to grayscale to help the model learn at a faster space[9].

In 2014, Goodfellow Ian introduced the term Generative Adversarial Networks (GAN) [10] and ever since then GAN has been useful as a solution for many problems. Since then, deep learning has been an additional help and the application of the combination has led to a massive evolution in the field of image processing. The development of Image super-resolution has been taken one step further when Dong [11] proposed the work of Super Resolution CNN. Over the years, image super-resolution performance has been increased exponentially. In our implementation, we also use Residual net [12] for obtaining the super resolution images and it helps us increase the depth of our sub-networks which helps us improve the image quality and hence solve the image enhancement problem of a surveillance system.

In this project, we use SRGAN to provide a solution to the problem of obtaining poor quality of images from the surveillance systems and hence the machine learning algorithm is utilized using a unique loss function for solving the poor image quality obtained from the CCTV system. It's framework consists of two sub-networks namely Generator

and Discriminator who compete with each other to provide us with enhanced images.

The remainder of this paper is organized as follows. Section 2 presents the related work and our contributions. Section 3 introduces our proposed SRGAN and section 4 is implementing the proposed model. Experimental results are presented in Section 5. Section 6 concludes the paper finally.

2. RELATED WORK

In this section, a survey analysis of the different proposed methods over the past few years for image enhancement of images obtained through surveillance systems (CCTV) is presented so that by analyzing the drawbacks of such methods we produce a new method that helps us utilize some previous methods in addition to some new methods to overcome all the present drawbacks which will provide us with optimal solution that gives us efficient result.

Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [13]

The paper proposes a super resolution generative adversarial network (SRGAN) to bring back all the minute details of an image and basically the images which lose their quality due to compression are being restored. Implementation of a loss function to the model has been presented. The perceptual loss consists of an adversarial loss and a content loss. The adversarial loss helps to obtain a more natural image using two sub-networks; generators and discriminators which are trained to differentiate between normal or down-sampled images and super resolution or upscaled images that are the original photo-realistic images. The content loss which is a euclidean distance between the reconstructed image to obtain better results. A dataset of BSD100 was used for training purposes and the PSNR value of the model was an average of 26.51 dB.

Disadvantages

The processed images have high peak SNR. The images lack high frequency texture and details. Standard quantitative measures such as PSNR and SSIM clearly fail to capture and accurately assess image quality. The proposed model is not optimized for Video SR in real time.

Removal of Noise Reduction for Image Processing. [14]
The paper proposes a solution for noise reduction using different types of filtering techniques. Noise is a result when an image is being acquired from some device due to which some pixel values get distorted and hence the image does not possess the ground truth value. Such methods can be used as a preprocessing measure for an input image obtained from a system. The different techniques are used for different noises. The grain noise in an image can be eliminated using linear filtering where each pixel gets set to average of pixels in their neighbourhood. The median filtering removes noise without decreasing the sharpness of the image by setting the output pixel to the median of neighbouring pixels of the input image, Adaptive filtering preserves all high frequency parts of an image due to its selective nature.

Disadvantages

The filtering techniques work quite good for preprocessing the images although the methods do not filter the images to a high extent. Among the three median filtering works with the best performance.

Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network [15]

Low resolution images are upscaled to high resolution using a filter and hence the super resolution must be applied to the high resolution space which increases the complexity. The paper proposes a Convolutional Neural Network which extracts feature maps from the low resolution image set. An array of upscaling factors is used to convert the low resolution images to high resolution. This helps in performing the two functions with less complexity and also using a higher version of upscaling factors. The super-resolution is kept at the end of the network and hence a sub-pixel convolution layer is used to upscale the image super-resolution. In addition the paper also proposes a deconvolution layer which recovers the resolution from down-sampling layer or from max-pooling.K2 Graphical Processing Unit (GPU) is used for performing the super-resolution of images. An average of 28.09dB PSNR is obtained from this model.

Disadvantages

The complexity of performing many tasks increases with many layers performing various tasks simultaneously. The super resolution is performed at the end which causes a slight delay.

Enhanced Deep Residual Networks for Single Image Super-Resolution[16]

The paper overcomes the disadvantage of complexity. The proposed model is a Multi-scale deep super resolution system(MDSR) which provides different high resolution images using different upscaling factors using a single model. The removal of different modules helps in optimizing the performance. Residual Networks have gained high popularity for converting low -level tasks to high-level tasks which

provide for low usage of the graphical processing unit. The residual scaling helps in stabilizing the training process. The result when upscaled with x4 sampling and training the model with B100 dataset is on an average 27.28dB. The multi-scale super resolution enables the reduction of model size and utilization of less time.

Disadvantages

The PSNR value is not optimized and the multi-scale remains compact while training on datasets.

3. PROPOSED MODEL

Surveillance systems require a model which enhances the images and serves the users of the systems with high positive experience. The model we propose is going to satisfy this goal. We propose a SRGAN model for enhancing images which are obtained from a CCTV camera.

A GAN is an adversarial network that has two sub-networks. The modules are the elementary foundation required for the system to perform the tasks in order to fulfill the objectives set for the project. There are two main modules used in the generative model used for unsupervised learning are the two sub-networks which are used to produce better clarity in the images. Here, the generative model captures the distribution of data and is trained in such a manner that it tries to maximize the probability of the Discriminator in making a mistake. The Discriminator, on the other hand, is based on a model that estimates the probability that the sample that it got is received from the training data and not from the Generator. The GANs are formulated as a minimax game that will compete with each other[10].

The GAN is used as a single architecture for denoising, super resolution and for a clear image. Hence this single architecture provides for various image processing tasks and helps us achieve better PSNR values of images we procure from a surveillance camera. The architecture of the proposed model is given at a detail in the subsection.

3.1 Architecture of the Proposed Model

The two networks help us produce images with lesser noise and better resolution. The generator network takes an image input of size 128x128 as shown in figure-1a with the rgb value set as 3 and hence we are taking all the three color parameters into consideration. We have added two up-scaling layers between the ResNets. The upscale layer does a 4x scale-up of the image resolution. The ResNets are the residual networks used for building multiple layers and connect the input layer to the output layer. A convolution neural network is used for sliding the filters over the image. It is used for filtering, segmenting and classifying the inputs. The filters used as a parameter in Residual Network also are a method to change the image in a certain way to provide a better result. Batch normalization is used as a catalyst for the training of the network. It helps us in reducing the number of epochs required by stabilizing the learning process. Finally we obtain the output of the generator using the tanh function. The image obtained as the output is upscaled to size 256x256. The tanh function squashes a real-valued number to the range [-1, 1]. Its output is zero-centered.

We train the SRGAN to learn how to perform the different tasks. We provide the down-sampled images and let the generator produce a real-like up-scaled image of the input. The discriminator if identifies the image as fake then the difference of the real image and the image produced by the generator is the noise which is used for the backpropagation and the generator and the discriminator learns from the tasks. This task is for the system to learn to produce better resolution images. For denoising, we feed images with high noise as inputs and for deraining we send images with the rain droplets on the images as input.

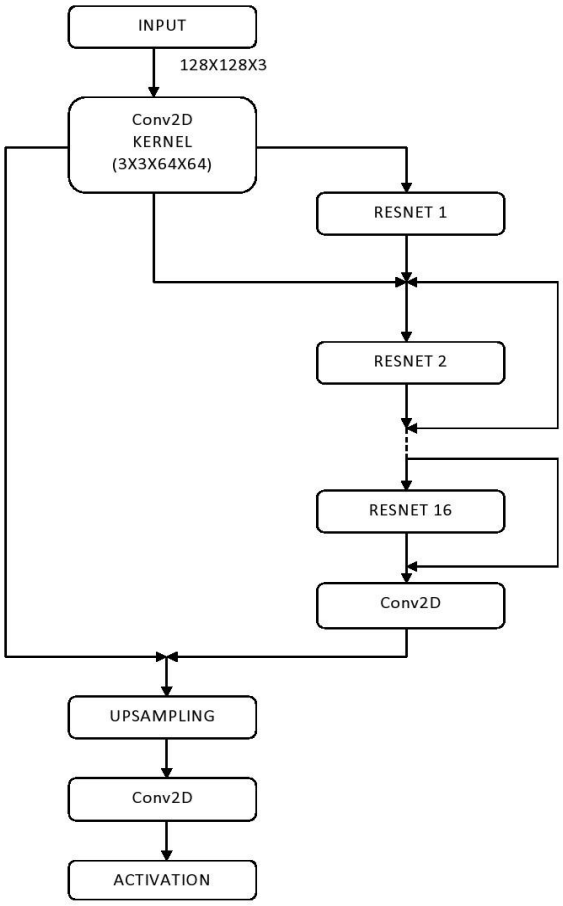

Figure 1a- Architecture of the generator

The output of the generator is fed to the discriminator network and hence the input to the discriminator is images of size 256x256 as shown in figure-1b. The convolution network is used as a linear non-saturating method to provide layers of input and finally a layer of output to discriminate between the real image and the fake image produced by the generator.

The rectified Linear Unit (ReLU) is used for helping with the activation to occur with threshold at zero. This greatly accelerates the noise function of stochastic gradient descent which is due to its linear, non-saturating features. The ReLU can be implemented in a simpler way than sigmoid and tanh by simply thresholding a matrix of activations at zero. The ReLU can die during training events since they are highly fragile and hence we use Leaky ReLU. Instead of the function being zero when x < 0, a leaky ReLU will instead have a small negative slope. The leaky ReLU helps to increase the range of the ReLU.

A dense layer is used in machine learning where every input is connected to every output. These layers have the output based on the units. The dense input layer has 1024 neurons and it provides the input to the next each neuron depending on the output of the previous neuron. The dense layer is used to implement the operation. The dense keyword is a core function in the keras library for machine learning and image processing. The batch normalization allows us to stabilize the inputs by standardizing them to have a mean of 0 and standard deviation as 1. The leaky ReLU is used as a method activation which is threshold at anything lesser than 0 or at 0.

The output for the discriminator uses the sigmoid function. The main reason why we use sigmoid function is because it exists between (0 to 1). Therefore, it is especially used for models where we have to predict the probability as an output. Since probability of anything exists only between the range of 0 and 1, sigmoid is the right choice. The probability is of whether it can detect if it is real or fake and this is used to calculate the error and is used for the backpropagation training of the generator and in turn also for the training of the discriminator.

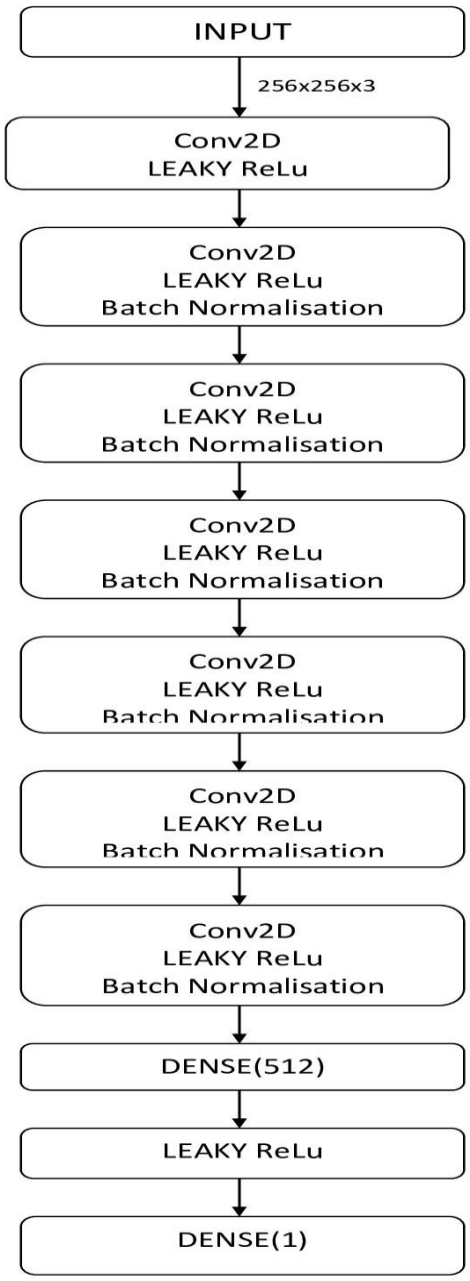

Figure 1b- Architecture of the Discriminator

4. IMPLEMENTATION

Implementation is the realization of an application, or execution of a plan, idea, model, design, specification, standard, algorithm or policy. The implementation of GAN takes place creating and training two artificial neural networks. The generator and discriminator are the two networks that are trained with each other using the minimax logic in the algorithm. The sum of the win and loss must be equal to one. We introduce a new refined loss function and

architectural novelties in the generator-discriminator pair for achieving improved results.

The loss function is aimed at reducing artifacts introduced by GANs and ensures better visual quality. We use two loss functions namely the Adam stochastic Gradient Descent and the Binary Cross Entropy Loss function for achieving the error value and using this value for the backpropagation algorithm and further training the model.

The basic steps used for constructing and predicting the module with the help of keras is:

1. Creating the network architecture with standard keras classes. Examples are Sequential, **Dense, Conv2D, Upsampling, BatchNormalisation.**
2. Compiling the created model using **model.compile ()** method.
3. Preprocessing the input dataset into tensors or by converting them into numpy arrays.
4. Preprocessing the target values for the dataset and converting them into tensors.

The first layer of the GAN, which takes a uniform noise distribution Z as input, could be called fully connected as it is just a matrix multiplication, but the result is reshaped into a 4-dimensional tensor and used as the start of the convolution stack. For the discriminator, the last convolution layer is flattened and then fed into a single sigmoid output.

Batch Normalization stabilizes learning by normalizing the input to each unit to have zero mean and unit variance. This helps deal with training problems that arise due to poor initialization and helps gradient flow in deeper models. This proved critical to get deep generators to begin learning, preventing the generator from collapsing all samples to a single point which is a common failure mode observed in GANs. Directly applying batchnorm to all layers however, resulted in sample oscillation and model instability. This was avoided by not applying batchnorm to the generator output layer and the discriminator input layer.

The ReLU activation is used in the generator with the exception of the output layer which uses the Tanh function. We observed that using a bounded activation allowed the model to learn more quickly to saturate and cover the color space of the training distribution. Within the discriminator we found the leaky rectified activation to work well, especially for higher resolution modeling. This is in contrast to the original GAN paper, which used the maxout activation (Goodfellow et al., 2013)[10]. For training the discriminator we have to produce some real labels (fake images) by the generator. Produce a batch of high resolution and low resolution images using the NumPy module and train the discriminator using the real images and the real labels.

All models were trained with mini-batch stochastic gradient descent (SGD) with a mini-batch size of 20. All weights were initialized from a zero-centered Normal distribution with standard deviation 0.02. In the LeakyReLU, the slope of the leak was set to 0.2 in all models. While previous GAN work has used momentum to accelerate training, we used the Adam optimizer with tuned hyperparameters. We found the suggested learning rate of 0.001, to be too high, using 0.0002 instead. Additionally, we found leaving the momentum term β1 at the suggested value of 0.9 resulted in training oscillation and instability while reducing it to 0.5 helped stabilize training.

5. RESULTS

A neural network contains hundreds of thousands of trainable parameters for which the gradient has to be computed in each epoch. The general implementation of a model is done using matrices on a dedicated GPU. The keras module uses a decent CPU or GPU for processing and executing the python converter file. Due to this factor Google's Colaboratory feature was used to train the model.

Google colab that allows you to start working directly on a **free Tesla K80 GPU** using Keras, Tensorflow and PyTorch, and how we can connect it to Google drive for the data hosting. It also gives you a total of 12 GB of ram, and you can use it up to 12 hours in row. Google Colab provides RAM of 12 GB with maximum extension of 25 GB and disk space of 358.27 GB.

Some of the observed results are depicted in figure-2 where the images are captured from an external CCTV camera and the images within the frames have been processed through the proposed model and the resolution has increased and thus provides a better quality image with reduction in the noise and hence better user experiences. The image provides a better vision and helps in many applications in the field of computer vision.

a)Input frame

b)Output Frame

c)Input frame

d)Output Frame

e)Input frame

f)Output Frame

Figure-2 Frames captured using surveillance system (a,c,e) and the enhanced version of the image in the frame using the proposed model (b,d,f)

The per pixel error is calculated by taking the absolute difference between the generated image and the target image.

The average accuracy percentage for multiple test cases are mentioned below

Dataset	Accuracy
Images from dataset	77.235%
Images outside the dataset	52.793%

Table1-Average accuracy percentage for different images

Time parameter for different operations	Value obtained from proposed model
Time for loading the model onto memory	15.296 sec
Time for normalization of dataset	113 ms
Time for a single prediction	3.522 sec
Total execution time	20.717 sec

Table-2 The average execution time for different processes

6. CONCLUSIONS

The paper proposes a SRGAN model for the image enhancement of the obtained inputs from productive surveillance systems. CCTV cameras have been an integral part of surveillance for some time now though the resolution that they shoot at is limited by the hardware used and the storage capacity available to them. A tool to produce a higher resolution image from a lower resolution image is all the more useful in this day and age. Although there are multiple ways of achieving this, using a popular technology to tackle this problem seems to be the way. A similar tool can be built using Generator adversarial Networks concepts from machine learning.

With the help of such a model, a logical connection can be achieved between a low resolution image and its higher version, which is learned by the model. The significant model provides an optimal solution to the problem of low resolution images of surveillance cameras than many other techniques available in the market.

It provides better accuracy and when compared with a measurement value such as PSNR, provides a much lesser noise in the images and hence less PSNR value with faster execution of the model and the implementation.Thus this is a step forward in producing images more pleasing to the eye at the least possible storage space required to store the image.

References

[1] Teague, C.; Green, L.; Leith, D. Watching me watching you: The use of CCTV to support safer workplaces for public transport transit offices. In Proceedings of Australian and New Zealand Communication Association Conference, Canberra, Australia, 9 July 2010

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

[3] N. N. A. N. Ghazali, N. A. Zamani, S. N. H. S. Abdullah and J. Jameson, "Super resolution combination methods for CCTV forensic interpretation," 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA), Kochi, 2012, pp. 853-858, doi: 10.1109/ISDA.2012.6416649.

[4] M. Sodanil and C. Intarat, "A Development of Image Enhancement for CCTV Images," 2015 5th International Conference on IT Convergence and Security (ICITCS), Kuala Lumpur, 2015, pp. 1-4, doi: 10.1109/ICITCS.2015.7292914.

[5] A. Chawdhary, S. Kumari, A. Bhavsar and R. Verma, "No Reference Evaluation in Super-Resolution for CCTV Footage," 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS), Rupnagar, India, 2018, pp. 107-112, doi: 10.1109/ICIINFS.2018.8721319.

[6] Jason Kurniawana , Sensa G.S. Syahraya , Chandra K. Dewab , Afia Hayati in Traffic Congestion Detection: Learning from CCTV Monitoring Images using Convolutional Neural Network, Jason Kurniawan et al. / Procedia Computer Science 144 (2018) 291–297

[7] Qiaojing Yan Stanford University Electrical Engineering,Wei Wang Stanford University Electrical Engineering ,"DCGANs for image super-resolution, denoising and deblurring", Published 2017.

[8] Nazare, Antonio & Ferreira, Renato & Schwartz, William. (2014). Scalable Feature Extraction for Visual Surveillance. 8827. 375-382. 10.1007/978-3-319-12568-8_46.

[9] Jaiswal, Varshali & Sharma, Varsha & Varma, Sunita. (2018). Comparative Analysis of CCTV Video Image Processing Techniques and Application: A Survey. 38-47.

[10] Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). Generative Adversarial Networks(PDF). Proceedings of the International Conference

on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680.
[11] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In European Conference on Computer Vision (ECCV), pages 184–199. Springer,2014
[12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition.In Washington, DC:IEEE Computer Society, pages770 — 778. 2016
[13] C. Ledig et al., "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, 2017, pp. 105-114, doi: 10.1109/CVPR.2017.19.
[14] Khaung Tin, Dr.Hlaing Htake. (2011). Removal of Noise Reduction for Image Processing.
[15] J. Kim, J. K. Lee and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 1646-1654, doi: 10.1109/CVPR.2016.182.
[16] B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, 2017, pp. 1132-1140, doi: 10.1109/CVPRW.2017.151.
[17] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel.Low-complexity single-image super-resolution based on nonnegative neighbor embedding. BMVC, 2012
[18] Kim, M., Park, D., Han, D. K., & Ko, H. (2014). A novel framework for extremely low-light video enhancement. InDigest of Technical Papers-IEEE InternationalConferenceon Consumer Electronics. (pp.91-92). [6775922] Institute ofElectrical and Electronics Engineers inc.10.1109/ICCE.2014
[19] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[20] K. Nasrollahi and T. B. Moeslund. Super-resolution: A comprehensive survey. In Machine Vision and Applications, volume 25, pages 1423–1468. 2014.

BIOGRAPHIES

Prof. M. Seshaiah, Received the M.Tech. degree in CSE from VTU, Belgaum where he is currently pursuing the Ph.D. His current research interests include digital image processing, compiler design and microprocessor & micro controllers.

Abhijith R Nair, he is an undergraduate student currently pursuing B.E CS&E in VTU Belgaum .His current research interests include digital image processing, machine learning.

Ektha Mallya, she is an undergraduate student currently pursuing B.E CS&E in VTU Belgaum .Her current research interests include digital image processing, machine learning.

Shiv Dev,, he is an undergraduate student currently pursuing B.E CS&E in VTU Belgaum .His current research interests include digital image processing, machine learning.