IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Augmenting of Corrupted Image from CCTV Footage

Dr.K.Geetha[1], R.Ragavi[2], M.Shanmuga priya[3], K.Swetha[4]

Dept. of Information Technology, M.A.M.College of Engineering and Technology, Triuchirappalli, India.

Abstract -Recently, image recognition technology using deeplearning has improved significantly. Forensic analysis has proved to be one of the most utilitarian tool in investigating crime. Forensic analysis provides evidence / basic informationof the said crime through analysis of physical evidence. This project address a scintillating technique to enhance the image quality of CCTV video to assist in the investigation of criminalcases. In the pre- processing phase for face recognition, face data with seven features that can be identified as a person are collected using CCTV. The collected dataset goes through the annotation process to classify the data and facial features are detected using deep learning. If there are four or more detectedfeatures, the image data is determined to be a person and the face is matched with stored user data in detail using 81 featurevectors. The problem of enhancing images is addressed by pureimage processing method and machine learning technique.This project analyzes both of the above techniques and further concluded that the machine learning approach produces a moreefficient result. The application of this technique can range from simple image filtering and advanced forensic image processing.

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

The Number of Closed-Circuit Television (CCTV) exponentially increases everyyear for several reasons. The reason could be due to the increasing crime rate, everything is put in a record which helps the crime department find the criminal based on evidence and recognition of the event or to regulate traffic violation [1]; hence the need for a surveillance system is always essential.

1. INTRODUCTION

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 to observe 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, and 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. The sub- images homomorphic filtering techniques is used to divide the image horizontally and vertically and enhancement is performed[4]; noise adaption superresolution 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 architecturewhich simplifies the model to high extent [6]; Deep Convolutional Generative Adversarial Networks allows us to simplify the architectureand helps us build a model which performs numerous image enhancement techniques such as image-super resolution, image- denoising and deconvolution that provides with optimum results for augment the images obtained from a CCTV system[7];Feature Extraction which is a computer vision technique and is well used for image filtering purposes [8]; hence is usually used as a preprocessing element.

For solving the limitations of the images procured from the CCTV, a few methods have proved to give a faster and systematic result. Images before being stuffed 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 key 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 grow the depth of our sub- networks which helps us upgrade the image quality and hence solve the image enhancementproblem of CCTV.

In this project, we use SRGAN to provide a solution to the problemof obtaining poor quality of images from the CCTV and hence themachine learning algorithm is utilized using a unique loss functionfor solving the poor image quality obtained from the CCTV system.Its framework consists of two sub-networks namely Generator and Discriminator who compete with each other to provide uswith 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 theproposed model. Experimental results are presented in Section
5. Section 6 concludes the paper finally.

2. RELATED WORK

In this section, a review of the different proposed methods overthe past few years for improvement of images obtained throughsurveillance systems (CCTV) is presented so that by analyzing the drawbacks of such methods that produce a new method thathelps to utilize some previous methods in addition to some newmethods 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 re-establish 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 adversarialloss 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 superresolution or exclusive images that are the original photo- realistic images. The content loss which is a Euclidean distance between the reconstructed images to obtain better solution. A dataset of BSD100 was used for training purposes and the PSNRvalue of the model was an average of 26.51 dB.

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 catch and exactly assess image quality. The proposed model is not improved 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 isbeing acquired from some device due to which some pixel values get warp 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 removed using linearfiltering where each pixel gets set to average of pixels in their neighborhood. The median filtering eliminates noise without decreasing the sharpness of the image by setting the output pixelto the median of neighboring pixels of the input image, Adaptive filtering preserves all high frequency parts of an image deto its selective nature. 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 anEfficient Sub-Pixel Convolutional Neural Network [15]

Low resolution images are up scaled to high resolution using a filterand 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 up scaling 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 up scaling factors. The super-resolution is kept at the end of the network and hence asub-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 forperforming the super-resolution of images. An average of28.09dBPSNR is obtained from this model. The complexity of performing many tasks increases with many layers performing various tasks simultaneously. The superresolution 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 exclusive factors using a single model. The removalof different modules helps in optimizing the performance. ResidualNetworks 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 stabilizingthe training process. The result when exclusive 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. The PSNR value is not developed and the multi-scaleremains compact while training on datasets.

Conclusion

In the existing system of extreme picture clarifier, takes a long time of processing the image. It supports only few type of images types. Zooming is not possible with this system. The system gives a poor resolution details of image even when effects are applied. Displacement of image has an improper view. The poor quality of the image mainly manifest in two aspects. One is the light attenuation and the counter measure should take is increasing the contrast, the other is color changing, need to white balance the original CCTV image to recover the distorted color to normal.

3. PROPOSED MODEL

Surveillance systems require a model which augment the imagesand serves the users of the systems with high positive experience. This model propose is going to satisfy this goal. To propose a SRGAN model for augmenting images which are obtained froma 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 probabilityof the Discriminator in making a mistake. The Discriminator, onthe other hand, is based on a model that estimates the probabilitythat the sample that it got is received from the training data and not from the Generator. The GANs are formulated as a minimax gamethat will compete with each other [10].

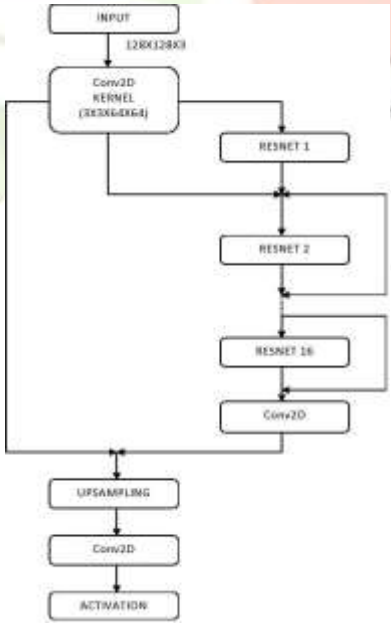
The GAN is used as a single architecture for denoising, super resolution and for a clear image. Hence this single architecture provides for varieties of image processing tasks and helps us to achieve better PSNR values of images that procure from asurveillance camera. The architecture of the proposed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

3.1 Architecture of the Proposed Model

The two networks help us to 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. Network have added two up-scaling layers between the ResNets. The upscale layer does a 4x scale-up of theimage resolution. The ResNets are the residual networks used forbuilding multiple layers and connect the input layer to the outputlayer. A convolution neural network is used for sliding the filtersover the image.

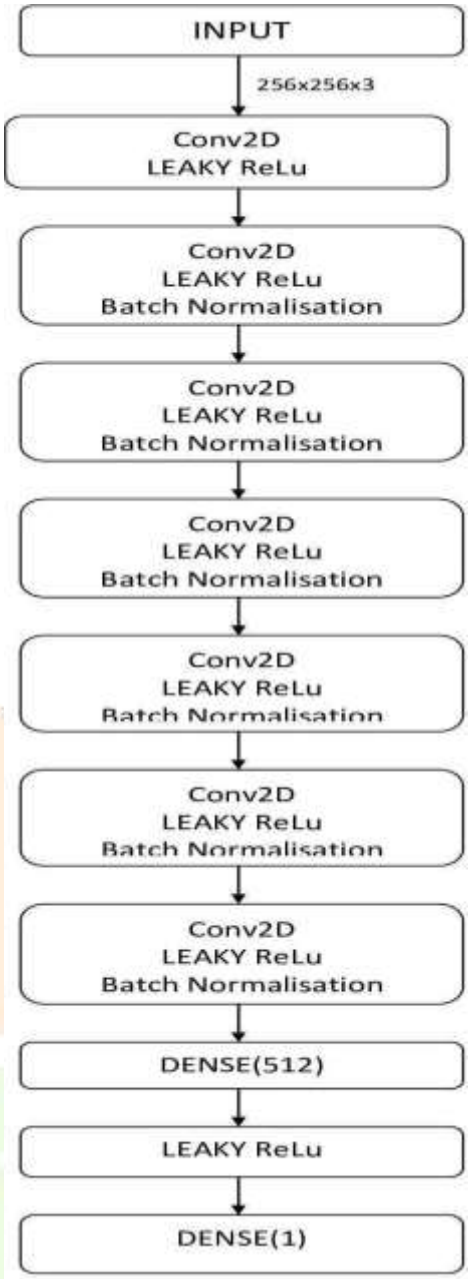
It is used for filtering, segmenting and classifying the inputs. The filters used as a parameter in Residual Network also are a method to change the image in a certain way to provide a better result. Batch normalization is used as a catalyst for the training of the network. It helps us in reducing the number of epochs requiredby stabilizing the learning process. Finally we obtain the output of the generator using the tanh function. The image obtained as theoutput is up scaled to size 256x256. The tanh function squashesa real-valued number to the range [-1, 1]. Its output is zero- centered.

This project train the SRGAN to learn how to perform the differenttasks. This provide the down-sampled images and let the generator produce a real like upscaled image of the input. The discriminatorif identifies the image as fake then the difference of the real imageand the image produced by the generator is the noise which is usedfor the back propagation and the generator and the discriminator learns from the tasks. This task is for the system to learn to producebetter resolution images. For denoising, we feed images with highnoise as inputs and for deraining we send images with the rain droplets on the images as input.



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 usedas 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 onthe units. The dense input layer has 1024 neurons and it providesthe input to the next each neuron depending on the output of theprevious neuron. The dense layer is used to implement theoperation. 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 leakyReLU 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 candetect if it is real or fake and this is used to calculate the error and is used for the back propagation training of the generator andin turn also for the training of the discriminator. 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 andfurther training the model.

The basic steps used for constructing and predicting the modulewith 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 byconverting them into numpy arrays.
4. Preprocessing the target values for the dataset andconverting 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 inputto each unit to have zero mean and unit variance. This helps dealwith training problems that arise due to poor initialization and helps gradient flow in deeper models. This proved critical to getdeep generators to begin learning, preventing the generator fromcollapsing all samples to a single point which is a common failure mode observed in GANs. Directly applying batchnorm toall 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 exceptionof the output layer which uses the Tanh function. We observed that using a bounded activation allowed the model to learn morequickly 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 theoriginal 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 usingthe 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 standarddeviation 0.02. In the Leaky ReLU, 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 rateof 0.001, to be too high, using 0.0002 instead. Additionally, we found leaving the momentum term $\beta 1$ 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 usingmatrices on a dedicated GPU. The keras module uses a decent CPUor GPU for processing and executing the python converter file. Dueto 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, Tensor flow and PyTorch, and how we can connect it to Google drive for the data hosting. It also givesyou 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 maximumextension 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 modeland 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 applicationsin the field of computer vision.

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 arementioned below;

INPUT FRAME

OUTPUT FRAME

INPUT FRAME

OUTPUT FRAME

INPUT FRAME

OUTPUT FRAME

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 higherresolution image from a lower resolution image is all the moreuseful in this day and age. Although there are multiple ways ofachieving this, using a popular technology to tackle this problem seems to be the way. A similar tool can be built usingGenerator 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. The use of CCTV to support safer workplaces for public transport transit offices.
[2] Yunbo Rao, Leiting Chen A Survey of Video Enhancement Techniques".
[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".
[4] M. Sodanil and C. Intarat, "A Development of ImageEnhancement for CCTV Images,"
[5] A.Chawdhary,S.Kumari,A.Bhavsar and R.Verma,"No Reference Evaluation in Super Resolution for CCTV Footage".
[6] Jason Kurniawaa, Sensa G.S.Syahraya, Chandra K.Dewab. Learning from CCTV monitoring image using convolutional neural network.
[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] Jalisa, 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. Proceedings of the International Conference.
[11] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In EuropeanConference on Computer Vision (ECCV), pages 184–199. Springer, 2014.
[12] HE K, ZHANG X, REN . Deep residual learning forimage recognition.
[13] C.Ledig" Photo-Realistic Single Image Super-Resolution using a Generative Adversial Network."
[14] Khaung tin, Dr.Hlaing Htake (2011).Removal of noise reduction.
[15] J. Kim, J. K. Lee and K. M. Lee, "Accurate Image Super- Resolution Using Very Deep Convolutional Networks," 2016IEEE Conference on Computer Vision and Pattern Recognition. [16] B. Lim, S. Son, H. Kim, S. Nah and K. M. Lee, "Enhanced DeepResidual Networks for Single Image Super- Resolution," 2017 IEEE Conference on Computer Vision and Pattern Recognition [17] Kim, M., Park, D., Han(2014). A novel framework for extremely low-light video enhancement.
[18] K. Nasrollahi and T. B. Moeslund. Super-resolution: A comprehensive survey.