

Suitability of Single Image Super-Resolution Models for Video Super-Resolution

Shreyas D G, V Sumanth Kaushik, Gururaja H S

Abstract: This project is an attempt to understand the suitability of the Single image super resolution models to video super resolution. Super Resolution refers to the process of enhancing the quality of low resolution images and video. Single image super resolution algorithms refer to those algorithms that can be applied on a single image to enhance its resolution. Whereas, video super resolution algorithms are those algorithms that are applied to a sequence of frames/images that constitute a video to enhance its resolution. In this paper we determine whether single image super resolution models can be applied to videos as well. When images are simply resized in Open CV, the traditional methods such as Interpolation are used which approximate the values of new pixels based on nearby pixel values which leave much to be desired in terms of visual quality, as the details (e.g. sharp edges) are often not preserved. We use deep learning techniques such as GANs (Generative Adversarial Networks) to train a model to output high resolution images from low resolution images. In this paper we analyse suitability of SRGAN and EDSR network architectures which are widely used and are popular for single image super resolution problem. We quantify the performance of these models, provide a method to evaluate and compare the models. We further draw a conclusion on the suitability and extent to which these models may be used for video super resolution. If found suitable this can have huge impact including but not limited to video compression, embedded models in end devices to enhance video output quality.

Keywords: EDSR, Image Enhancement, Super Resolution, SRGAN, Video Enhancement

I. INTRODUCTION

On new year’s eve in 2018, WhatsApp servers had crashed globally and no one could send or receive any messages. The large number of GIFs shared by the users during Christmas and new year overloaded the servers. Most of the televisions these days feature 4K and 8K screens, but the TV channels and streaming services do not provide a resolution of more than 1080p. One possible solution to these problems is to use the concept of super resolution. In the first case we can reduce the resolution of the shared media and later enhance it in the user’s end device. For the later problem, super resolution models could be built into televisions and this could provide for a better viewing experience. Super resolution refers to the process of enhancing the details and resolution of an image. Often a low quality image is taken as input and the same image is enhanced to a higher resolution, which is later given out as output.

Revised Manuscript Received on May 25, 2020.

* Correspondence Author

Shreyas D G*, Student, Department of Information Science and Engineering, BMS College of Engineering affiliated to Visvesvaraya Technological University, Bangalore, India.

V Sumanth Kaushik, Student, Department of Information Science and Engineering, BMS College of Engineering affiliated to Visvesvaraya Technological University, Bangalore, India.

Gururaja H.S., Assistant Professor in the Department of Information Science and Engineering, BMS College of Engineering (BMSCE), Bangalore, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The challenge here is to fill details which are essentially unknown in the low resolution image. Most super-resolution techniques are based on the same idea: using information from several different low resolution images to create one enhanced image.

II. RELATED WORK

Super resolution received substantial attention from within the computer vision research community and several approaches have been attempted to achieve this task.



Figure 1. Super resolution on a lower resolution image of human face.

Super-resolution GAN applies a deep network in combination with an adversary network to deliver higher quality pictures[1]. A human will find an SRGAN more appealing with more details in contrast to the structure of GAN without SRResNet. During the training, A high-resolution image (HR) is down-sampled to a low-resolution image (LR). A GAN generator up-samples LR pictures to super-resolution pictures (SR). We utilize a discriminator to differentiate the HR images and backpropagate the GAN loss to train the Discriminator and the generator[3]. An enhanced deep super-resolution network (EDSR) exceeds the performance of current state of the art Super Resolution methods. The significant performance improvement of the EDSR models is due to optimization by removing unnecessary modules in conventional residual networks[2]. The performance is further improved by expanding the model size stabilizing the training procedure. For video super resolution traditional algorithm based approaches which involve interpolation techniques and inference of pixel intensities from previous frames have been used.

III. PROPOSED APPROACH

As shown in figure 2, a low resolution image is passed through a normalizer which normalizes the image, does some basic pre-processing and passes it to the Super-Resolver which is an SRGAN or EDSR model.

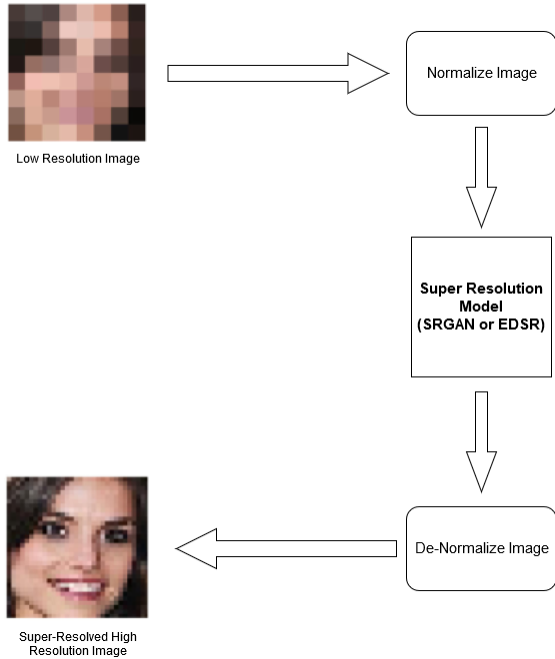


Figure 2. High Level Diagram

The model outputs a super-resolved image which is de-normalized and displayed as the output of the model.

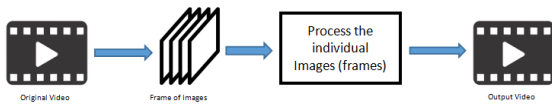


Figure 3. Extending the model to support videos

The model is then extended to videos by breaking the videos into individual frames applying the super-resolution model to each of these individual frames and stitching back the video from these super-resolved/enhanced frames. This process is shown in Figure 3. Then the suitability of applying the above mentioned single image models for videos is checked.

Later a comparative analysis of the SRGAN model that was built with an already trained EDSR model is performed to compare the two models in terms of performance, processing time, processing power required, efficiency, etc.

IV. EVALUATION

As shown in the low level data flow diagram we first select a low resolution video whose quality is to be increased and the corresponding high resolution video which serves as the ground truth. We then pass these

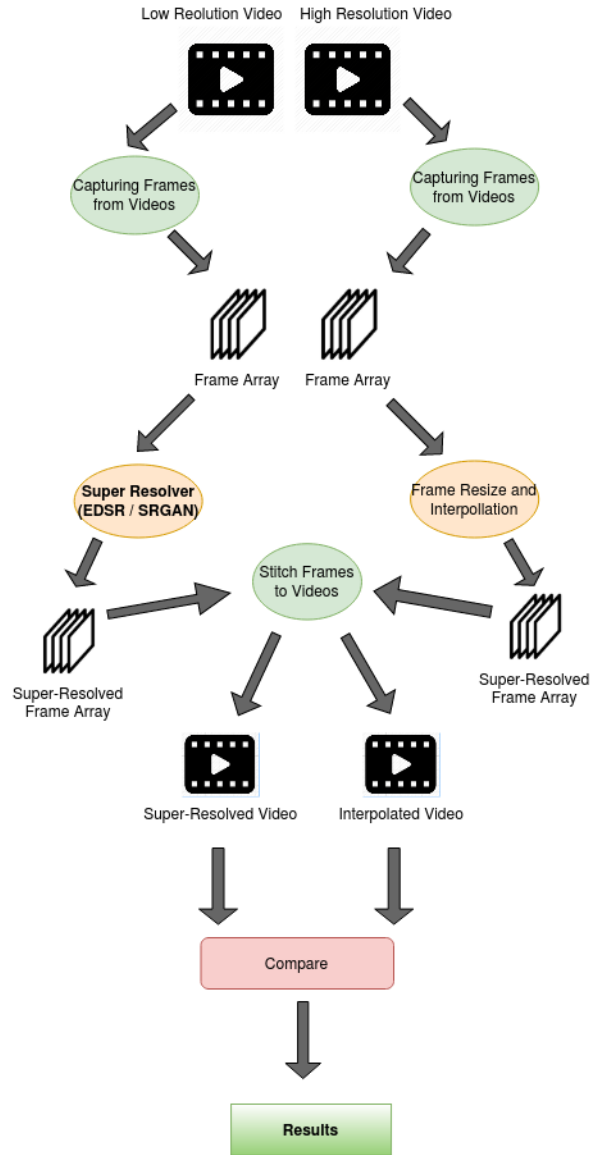


Figure 4. Low Level Diagram

two videos through a frame captor which captures the individual frames from both the videos and stores the frames as images in separate directories. Then these individual frames from the low resolution video is passed to the Super-Resolver which is the SRGAN and EDSR models for enhancing the resolution of an image. The corresponding frames of the high resolution video is passed into a Frame Re-sizer which uses interpolation to change the frame size. Interpolation is done because a video player would interpolate the frames of a video in order to adjust to the size of the screen when the video is being played. For the purpose of comparing the output of the Super-resolver and the high resolution frames both need to be of the same size. The purpose of this comparison is to check how close is the super-resolved image to that of ground truth high resolution frames. The modified frames are passed to a comparing block which extensively performs several comparisons and quantifies the difference between the super-resolved images and resized ground truth frames.

The performance of the super-resolved videos obtained by using the SRGAN and EDSR models is also compared here. The frames are then stitched back into videos. Now we have two videos one which is the Super-Resolved video whose resolution is increased using the SRGAN or EDSR model and the other is the interpolated high resolution video. The final results of the comparisons are presented as graphs.

V. EVALUATION METRICS

A. Average pixel loss :

Average pixel loss is computed by taking the output image of the model and comparing with the ground truth image. The corresponding Pixel intensity difference is computed for each pixel and the average loss is calculated.

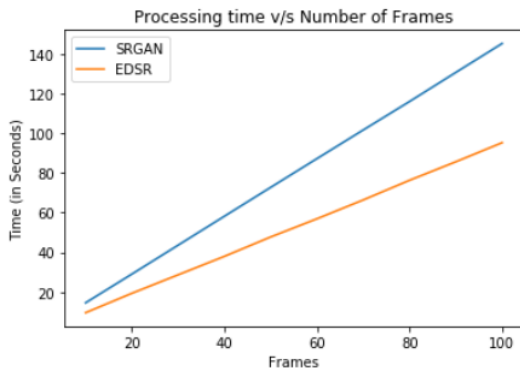
B. Manhattan distance loss :

Manhattan distance loss is computed by taking the output image of the model and comparing with the ground truth image. Each pixel is considered and the Manhattan distance is computed for corresponding pixel intensities.

C. Euclidian distance loss :

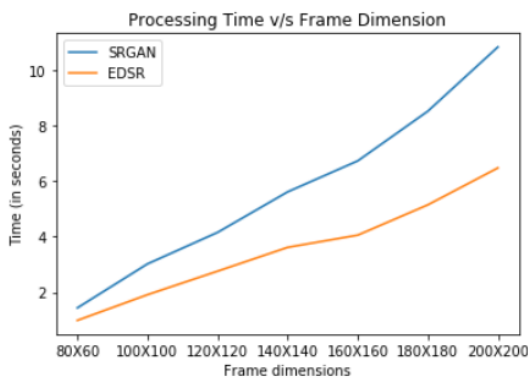
Euclidian distance loss is computed by taking the output image of the model and comparing with the ground truth image. Each pixel is considered and the Euclidian distance is computed for corresponding pixel intensities.

VI. RESULTS



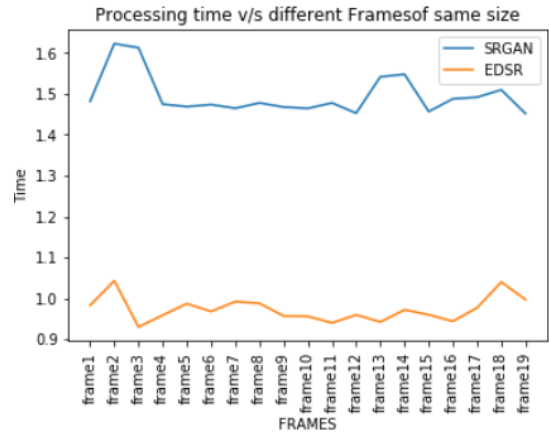
Aim: To Visualize the relationship between processing time and no of frames.

Inference: Processing time increases linearly with number of frames. The processing time for the EDSR model is less than the SRGAN.



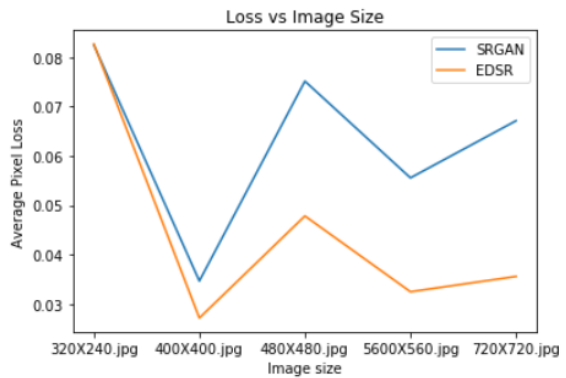
Aim: To visualize the relationship between processing time and frame dimensions.

Inference: Processing time increases non-linearly with frame size. The processing time for the EDSR model is less than the SRGAN for all frame sizes.



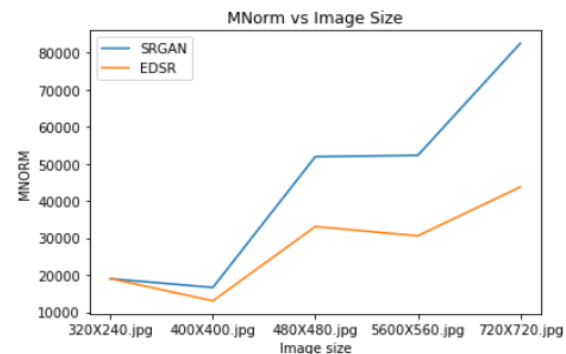
Aim: To visualize the effect of different frames/parts of the same image on processing time.

Inference: The variation in processing time about the mean processing time for both models is not much. The processing time for the EDSR model is less than the SRGAN for all frames.



Aim: To visualize how the average pixel loss is affected by image size.

Inference: Average pixel loss although there is slight variation in general it tends to decrease with increase in size of image.

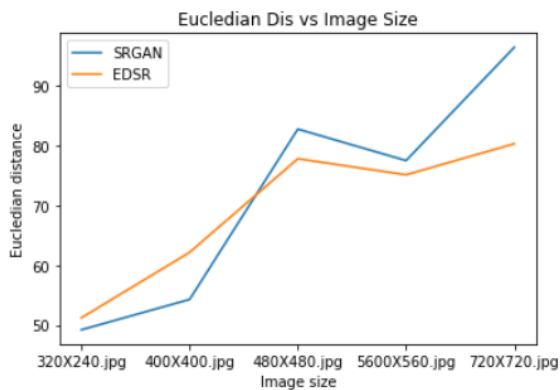


Aim: To visualize how the Manhattan distance is affected by image size.

Inference: Increase in Manhattan distance is non-linear with increase in image size.

Suitability of Single Image Super-Resolution Models for Video Super-Resolution

The distance is less for the EDSR model as compared to SRGAN model.



Aim: To visualize how the Euclidian distance is affected by image size.

Inference: Increase in Manhattan distance is non-linear with increase in image size. Distance is less for the EDSR model as compared to SRGAN model.

VII. CONCLUSION

The EDSR model performs better than the SRGAN model. This is confirmed by all the six graphs show in the results. In terms of Processing time the EDSR model takes less time than the SRGAN model in all cases without any exception. This shows that the EDSR model is less processing power than the SRGAN model.

The Accuracy of the super resolution is very high. This is confirmed by the average pixel loss whose value is very less in the order of 0.04.

The super resolution model is easily applicable for GIFs(Graphic Interchange Format). The frame rate of a GIF is usually low around 10 to 15 frames per second. Even the duration is very small, less than 5 seconds with most of them being less than 3 seconds. The processing time taken for enhancing the resolution of an entire GIF is less than 4 seconds and that is without a GPU. A GPU enabled server can do this enhancement in a fraction of a second.

The processing power required to enhance the quality of a high resolution video to an even higher resolution is very high. Even with a powerful PC it will take very long to enhance the quality of high resolution video. As of now processing is the main bottle neck for the implementation of the concept of super resolution for practical use such as enhancing the videos in 4k and 8k TVs.

From these experiments we can conclude that single image super resolution models are not suitable for video super resolution in end devices with low processing capabilities. However, in environments with high processing capabilities and no real-time requirements (situations where high latency is acceptable) it can be used.

REFERENCES

1. Christian Ledig, Lucas Theis, Photo-Realistic SingleImage Super-Resolution_Using a Generative Adversarial Network, IEEE, 25 May 2017.
2. Bee Lim, Sanghyun Son , Enhanced Deep Residual Networks for Single Image Super-Resolution, IEEE, 10 Jul 2017.

3. Kaiming He, Xiangyu Zhang , Deep Residual Learning for Image Recognition , IEEE , 10 Dec 2015.
4. <https://pathmind.com/wiki/neural-network>
5. https://en.wikipedia.org/wiki/Neural_network
6. https://en.wikipedia.org/wiki/Artificial_neural_network
7. <https://towardsdatascience.com/understanding-neural-networks-19020b75823>
8. <https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414>
9. <https://www.geeksforgeeks.org/neural-networks-a-beginners-guide/>
10. <https://www.geeksforgeeks.org/generative-adversarial-network-gan/>
11. <https://pathmind.com/wiki/generative-adversarial-network-gan>
12. <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/>
13. <https://towardsdatascience.com/understanding-generative-adversarial-networks-gans-cd6e4651a29>
14. https://en.wikipedia.org/wiki/Generative_adversarial_network
15. https://medium.com/@jonathan_hui/gan-whats-generative-adversarial-networks-and-its-application-f39ed278ef09
16. <https://developers.google.com/machine-learning/gan>

AUTHORS PROFILE

Shreyas D G currently pursuing final year of B. Tech in Information Science and Engineering in B.M.S. College of Engineering, Bangalore affiliated to Visvesvaraya Technological University. Has had research experience at Cisco Systems, India and CIIRC. Has attended multiple workshops on robotics and IoT from IIT Kharagpur, Techie nest robotics workshop etc. Is also a certified technology associate in Microsoft Azure. Has completed multiple certifications on machine learning, deep learning and bioinformatics from coursera. Has done considerable research and projects in ML. Keen on exploring the different applications of ML and DL that can help humanity.



V Sumanth Kaushik Currently a final year undergraduate student in dept. Of information science and engineering at BMS College of Engineering affiliated to Visvesvaraya Technological University. Have had research experiences at IISc and VMware Inc. Has worked on multiple interdisciplinary research projects. Has attended multiple workshops on robotics from IIT Kharagpur, Techie nest robotics workshop etc. Is also a certified technology associate in Microsoft Azure. Has completed multiple certifications on machine learning, deep learning, robotics and data science from coursera, udemy, edx. Have also received multiple scholarships from BMSCE Alumni association, Huawei Inc. for academic performance . Broadly interested in computer vision, robotics and artificial intelligence. Has a publication in the field of computer vision and machine learning in an International Conference.



Gururaja H.S. obtained his Bachelor of Engineering (B.E.) in Computer Science & Engineering and Master of Technology (M. Tech) in Computer Network Engineering from VTU, Belgaum. He is pursuing his doctoral research in the area of Network Security from JNTU, Hyderabad. He has around 14 years of teaching experience and has been awarded as Best Teacher in his previous workplace. Currently he is working as an Assistant Professor in the Department of Information Science and Engineering, BMS College of Engineering (BMSCE), Bangalore. He has more than 10 publications in International / National Scopus/SCI indexed journals and conferences to his credit. He has been certified as a top performer for Mentoring Educators in Educational Technology by IIT-Bombay. He has given webinars and presented papers on learning methodologies in Higher Education. His area of research includes Network Security and Cryptography, Machine Learning and Robotic Process Automation.

