ORIGINAL RESEARCH PAPER
Volume-9 | Issue-3 | March-2020 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr
INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

ANALYSIS OF ENHANCEMENT OF COMPRESSED VIDEO QUALITY USING GENERATIVE ADVERSARIAL NETWORKS (GAN): A SURVEY

Computer Science

Koustav Guha B.Tech Student, Department Of Computer Science And Engineering, University Of Engineering And Management Kolkata

Soumyajit Chakraborty* B.Tech Student, Department Of Computer Science And Engineering, University Of Engineering And Management Kolkata *Corresponding Author

ABSTRACT

In commercial video streaming services, the video compression uses lossy algorithms in such a manner it lowers the bandwidth while transmitting them. As we know GAN can obtain very high quality images in the enhancement tasks of images but it involves large generators that are very costly and also takes a lot of processing time. Our paper will provide an architecture that will do the same task at a much lower price.

KEYWORDS

High Resolution, Low Resolution, Single Image Super Resolution, Generative Adversarial Networks, Artifacts , Convolutional Neural Networks

1. INTRODUCTION :

Daily huge number of videos are produced, streamed and shared on the web, and lots are produced, used within private systems like cameras and mobile phones. So, to store and to transmit these video streams it is very important to compress them, to decrease the storage requirements and bandwidth. Basically video is compressed using lossy algorithms, especially when we are dealing with HD and 4K resolution videos. The result of these algorithms causes more or less strong loss of content fidelity in comparison to the original visual data, to get a better compression ratio. Since one of the factors that accounts for user experience is image quality, compression algorithms are designed to reduce perceptual quality loss.

Another example that we use in our day to day lives where a high compression is desirable is that of video conferencing, where video streams must be kept small to reduce the communication latency and thus improve user experience. One of the most use case is that of videos encoded and transmitted from IoT devices and drones, in which there is the low-energy constraint that calls for high compression of the streams to reduce transmission time, to save battery power.

While compressing the videos many artifacts appear. These artifacts are because of different kinds of lossy compressions used.

In this method we have proposed a solution to the artifact removal which is based on Convolutional Neural Networks that have been trained on a large sets of frame patches compressed with different quality factors. Our approach is independent with respect to the compression algorithm used to process a video. We can apply these methods on lots of lossy compression algorithms like WebM, AV1, H.264/AVC and H.265/HEVC.

One of the main merit of enhancing video quality working on artifact removal of our method is that it can be applied just on the receiving end of the coding pipeline, e.g. using dedicated hardware such as SoCs or GPUs.

2 RELATED WORK:

Improving image quality is a topic that has been studied a lot from the past , especially if we consider the case of compression artifact removal , most of them are based on the techniques of image proce ssing.

Fig.A : Examples of video compression artifacts: top to bottom – (top)Staircase noise (Spatial), (middle left) Mosquito noise (Temporal), (middle right) Blurring (Spatial), (bottom) Blocking (Spatial)

Best results are obtained using DCNN Deep (Convolutional Neural Networks), trained to restore image quality using couples of undistorted and distorted images. Dong et al. extended their previous work on super-resolution SRCNN with artifact removal CNN (AR-CNN) sharing a common architecture with SRCNN, following sparse coding pipelines.

2.1 CONTRIBUTION

All the applications for such technology in video quality enhancement have a less or more strict real-time constraint. In this paper we are proposing a GAN in which the generator is designed with efficiency in mind. In place of that, the discriminator can be made large at will, as it only affects training efficiency. We show that on no-reference video and image quality assessment our method makes frames that have scores higher than compressed frames.

3 METHODOLOGY :

Considering a raw frame It from a sequence we consider $I^C_t = C(I_t, I_{t-1}, I_{t-2}, \ldots, \theta)$ as its compressed version, where $C(\cdot)$ is some compression algorithm for video sequences such as H.264/AVC configured according to a parameters set θ.

If we represent the images as real valued tensors in R^{HxWxCh}, where H and W are height and width of the frame, and Ch is the number of channels, function $G(\cdot)$ able to invert the compression process:

$$G(I^C_t) = I^R_t \approx I_t \quad (1)$$

here I^R_t denotes the restored version of I^C_t . In all models we use Ch = 3, as we are training over and restoring RGB video frame. The function $G(\cdot)$ as a fully convolutional neural network.

Training of "fake" images is able to induce the discriminator in mistakes. GAN consists of the optimization of two networks named generator (G) and discriminator (D) where the generator is fed some noisy input and has the goal to create "fake" images to be able to induce the discriminator in mistakes. Moreover the discriminator optimizes the classification loss rewarding solutions that correctly distinguish fake from real images. Our task regards the enhancement of a corrupted image. Here, our end goal is to obtain a $G(\cdot)$ function that

will be able to process compressed frames and remove the artifacts. In our conditional GAN we provide to the discriminator positive examples and negative examples.

3.1 GENERATIVE NETWORK

The generator that we are using, its architecture is based on MobileNetV2, that is a very efficient network for cellular devices to perform classification tasks. If we replace standard residual blocks with bottleneck depth-separable convolutions blocks, as shown in Table 1.0, to reduce the total amount of parameters. We have set the expansion factor t to 6 for all the experiments.

Table 1.0: Bottleneck residual block used in our generator network

Layer	Output
Conv2d 1 × 1, ReLU6	m × n × t * c
Dw Conv2d 3 × 3, ReLU6	m × n × t * c
Conv2d 1 × 1	m × n × c

After first standard convolutional layer, we halved the feature maps twice and then we apply a chain of B bottleneck residual blocks. The number of convolution filters doubles each time the feature map dimensions are halved. We use two combinations of nearest-neighbour up-sampling and standard convolution layer to restore the original dimensions of feature maps. Finally, we generate the RGB image with a 1 × 1 convolution followed by a tanh activation.

3.2 DISCRIMINATIVE NETWORK

The discriminator network consists mainly of convolutional layers, i.e followed by LeakyReLU activation, with a sigmoid activation and a final dense layer. As the complexity of this network is not affecting the execution time during test phase, we have chosen for all our trainings a discriminator with a very large number of parameters.

3.3 CONTENT LOSSES

In this paper we have described the content loss used also with the adversarial loss for the generator. The content loss is a pixel-based reconstruction loss function. The L1-norm and L2-norm are used generally for Super Resolution. Aim of the content losses is to limit the set of distributions to be modelled via the adversarial learning process inducing the generator to produce consistent image enhancement behaviour.

PIXEL WISE MSE LOSS

This loss is generally used in image restoration tasks and image reconstruction. It recovers frequency details that are low from a compressed image, still the demerit is that high frequency details are suppressed.

PERCEPTUAL LOSS

Many contributions on image enhancement, restoration and super resolution have used the perceptual loss to optimize the network in a feature space rather than in the pixel space. In our adversarial training, we have used such loss to encourage reconstructed images and target ones to have the similar feature representations. By projecting I and I^R on a feature space of a pre-trained network, hence extracting some meaningful feature maps, the measure of similarity between two images is obtained

3.3 DIFFERENTIAL CONTENT LOSS:

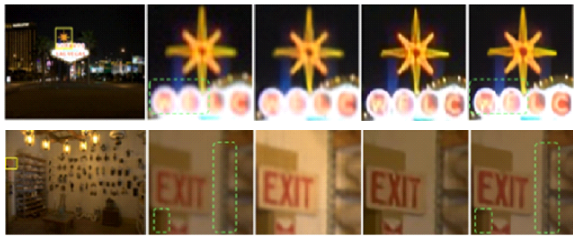
The differential content loss evaluates the difference between the SR and HR images in a much deeper level. It can help to decrease the over-smoothness and also in improving the performance of reconstruction especially for high frequency components

4 EXPERIMENTAL RESULTS:

We have tested our novel architectures with three popular no reference metrics. No reference image quality assessment is the task of providing a score for an image, that has been possibly distorted by unknown process, without having the access to original image. These metrics are designed to identify and quantify the presence of different kinds of artifacts that might be present in the image being analysed.

All models have been trained on DIV2K dataset. DIV2K training set consists of 800 uncompressed images of high resolution, that we have compressed using H.264 to generate degraded frames. As an augmentation strategy, if we consider the size of DIV2K which is small, we have resized the images at 512, 256 and 384 on its shorter side and then we have randomly cropped a patch of 224 × 224 pixels with a random mirror flipping. This step allows to increase dataset size and increase diversity of pattern scales.

Looking at examples in the given figure i.e at Fig **B**, it can be seen that the quality of imagery is largely improved by our networks in comparison with the source compressed frame. Large frame obtained by our Fast network.



(I) (ii) (iii) (iv) (v)

Fig.B Fine details of the texture of the water and feathers of the duck obtained by the GAN based approach, compared to the standard augmentation strategy.

So, GAN image enhancement can lead to low performance in full reference metrics due to the fact that the overall pictures is improved by semantically consistent textures which, pixel-wise, may differ from the original uncompressed image.

5 CONCLUSION

This paper provides a short survey of analysis of enhancement of Compressed Video Quality using GANs. Lots of applications based on video streaming impose a strict real-time constraint. The fast network is able to run at 20 FPS with no deterioration on the final results. Qualitative inspection of the frames confirms quantitative results, showing pretty highly detailed frames.

REFERENCES

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on Single Image Super-Resolution: Dataset and study. In: Proc. of IEEE CVPR Workshops (2017)
2. Cavigelli, L., Hager, P., Benini, L.: CAS-CNN: A deep convolutional neural network for image compression artifact suppression. In: Proc. of IJCNN (2017)
3. Chu, M., Xie, Y., Leal-Taixe, L., Thuerey, N.: Temporally coherent GANs for video super- ´ resolution (tecogan). arXiv preprint arXiv:1811.09393 (2018)
4. Dar, Y., Bruckstein, A.M., Elad, M., Giryes, R.: Postprocessing of compressed images via sequential denoising. IEEE Transactions on Image Processing 25(7), 3044–3058 (July 2016)
5. Dosovitskiy, A., Brox, T.: Generating images with perceptual similarity metrics based on deep networks. In: Proc. of NIPS (2016)
6. Galteri, L., Seidenari, L., Bertini, M., Del Bimbo, A.: Deep universal generative adversarial compression artifact removal. IEEE Transactions on Multimedia pp. 1–1 (2019)
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Proc. of NIPS (2014)
8. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 21(12), 4695–4708 (Dec 2012)
9. Leonardo Galteri, Lorenzo Seidenari, Marco Bertini .: Deep Universal Generative Adversarial Compression Artifact Removal , IEEE TRANSACTIONS ON MULTIMEDIA
10. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study
11. Jakhetiya, V., Lin, W., Jaiswal, S.P., Guntuku, S.C., Au, O.C.: Maximum a posterior and perceptually motivated reconstruction algorithm: A generic framework. IEEE Transactions on Multimedia 19(1), 93–106 (2017)
12. Kang, L.W., Hsu, C.C., Zhuang, B., Lin, C.W., Yeh, C.H.: Learning-based joint superresolution and deblocking for a highly compressed image. IEEE Transactions on Multimedia 17(7), 921–934 (2015)
13. Mittal, A., Saad, M.A., Bovik, A.C.: A completely blind video integrity oracle. IEEE Transactions on Image Processing 25(1), 289–300 (2016)
14. Mittal, A., Moorthy, A.K., Bovik, A.C.: No-reference image quality assessment in the spatial domain. IEEE Transactions on Image Processing 21(12), 4695–4708 (Dec 2012)
15. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proc. of ICLR (2015)