# Quality Enhancement of Gaming Content using Generative Adversarial Networks

Nasim J. Avanaki*, Saman Zadtootaghaj*, Nabajeet Barman†, Steven Schmidt *,
Maria G. Martini†, Sebastian Möller*

*Quality and Usability Lab, TU Berlin, Germany,

jamshidiavanaki@campus.tu-berlin.de, saman.zadtootaghaj@qu.tu-berlin.de, {steven.schmidt, sebastian.moeller}@tu-berlin.de

†Wireless Multimedia & Networking Research Group, Kingston University, London, UK,

{n.barman, m.martini}@kingston.ac.uk

*Abstract*—**Recently, streaming of gameplay scenes has gained much attention, as evident with the rise of platforms such as Twitch.tv and Facebook Gaming. These streaming services have to deal with many challenges due to the low quality of source materials caused by client devices, network limitations such as bandwidth and packet loss, as well as low delay requirements. Spatial video artifact such as blockiness and blurriness as a result of as video compression or up-scaling algorithms can significantly impact the Quality of Experience of end-users of passive gaming video streaming applications. In this paper, we investigate solutions to enhance the video quality of compressed gaming content. Recently, several super-resolution enhancement techniques using Generative Adversarial Network (e.g., SRGAN) have been proposed, which are shown to work with high accuracy on non-gaming content. Towards this end, we improved the SRGAN by adding a modified loss function as well as changing the generator network such as layer levels and skip connections to improve the flow of information in the network, which is shown to improve the perceived quality significantly. In addition, we present a performance evaluation of improved SRGAN for the enhancement of frame quality caused by compression and rescaling artifacts for gaming content encoded in multiple resolution-bitrate pairs.**

## I. INTRODUCTION

Video traffic forms a significant part (almost 82% by 2021 as per current Cisco Visual Networking Index forecast [1]) of the net consumer Internet traffic. Along with the rising popularity of video-on-demand streaming services such as Netflix and YouTube, live streaming services such as Twitch and Facebook Live have got increasing acceptance by the users. Among the live streaming services, gaming video streaming services which broadcast live gameplay to viewers as provided by Twitch and Facebook Gaming have become hugely popular with Twitch alone consisting of almost 15 million daily active users and 2 million streamers resulting it in being the $4^{th}$ highest peak traffic generator in the US. One of the measures to adapt the video content to the user bandwidth is by compressing and/or rescaling the video into multiple resolution-bitrate pairs, as is done by almost all major Over The Top (OTT) service providers. The selection of the resolution-bitrate pairs is determined based on a trade-off between blockiness (due to compression) and blurriness (due to rescaling), as is explained in detail in [2].

In the past few years, we have seen several works to build an enhancement technique that can scale up an image to a higher resolution without significant quality loss, known as the Super-Resolution task (SR). Recently, there has been a huge improvement in the performance of the methods for SR task, primarily due to two big factors: advancement in deep learning methods (such as deep Convolutional Neural Networks (CNN)), and improvement in loss function to evaluate the quality of the enhanced image. Several new CNN architectures have been introduced that improve the prediction of different tasks by either proposing a deeper network (e.g., VGG [3]) or enhancing the flow of information (e.g., DenseNet [4] and ResNet [5]). Besides, new and more versatile machine learning methods such as Generative Adversarial Networks (GANs) [6] and Auto-encoders have found their application in SR tasks [7]. GANs, which have two main blocks, generator and discriminator has recently gained lots of attention for quality enhancement, especially for SR tasks. In SR tasks, the generator is responsible for generating fake images based on the received distorted input, while discriminator evaluates the generated image quality compared to the real image. GANs have been considered as a solution for quality enhancement of distorted images due to compression, noise, or block loss [8].

Recent studies in [9], [10], and [11] have shown that gaming content due to its synthetic and artificial nature, is different from non-gaming content and it is imperative that models and techniques developed for non-gaming content needs to be adapted for gaming content for increased performance efficiency. A game is created by a pool of pre-designed objects that are repeated several times in different game scenarios. Such repetition of content structures might help train a deep neural network which can predict the quality or perform other similar tasks with very high accuracy due to similarity in content structure between the training and test dataset. Such a similarity has been proved to play an important role in the very good performance of GANs in SR tasks [12]. Given the fact that almost 50% of streamed gaming videos come from the top 10 highly popular games there is a vast potential of applicability of game-specific quality enhancement/assessment models in real-world applications.

Towards this end, in this paper, we first investigate the performance of GANs for quality enhancement of encoded frames of video games. Figure 1 illustrates the architecture of SRGAN for quality enhancement. We first evaluate the enhancement
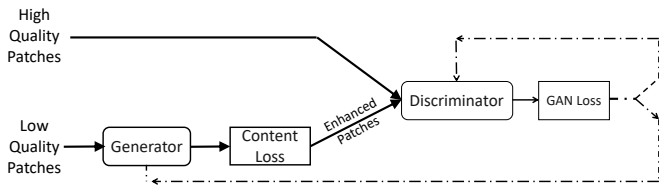
Fig. 1: The architecture of SRGAN for quality enhancement.

model SRGAN [13] for gaming image enhancement, followed by proposal of a new generative network architecture which improves the flow of information as well as improving the loss function of SRGAN by allowing the content loss function of SRGAN to learn quality prediction task.

The rest of the paper is organized as follows. Section II presents some of the related work in the field of image quality enhancement. Section III presents the different datasets used for training and test and Section IV presents the model development steps for the proposed model. Section V presents the results of the objective and subjective tests of the proposed model. Section VI concludes the paper.

## II. RELATED WORK

In general, the proposed methods in the field of image compression artifacts reduction can be divided into two main categories: deblocking-oriented and deep learning-based approaches. The first category of approaches aim to eliminate ringing and blockiness artifacts. However, the weakness of these methods is that, while they can manage to eliminate the blocking artifacts, the edges of the original image can not be as sharp as they originally were. The second category of approaches have recently caught attention due to the advancement of deep learning methods. Several enhancement models are proposed and applied on compressed images using deep learning methods such CNNs. [14] is one of the very first works that used Convolutional Neural Network (CNN) for quality enhancement of images with compression artifacts. In another work, Zhang et al. proposed a feed-forward denoising convolutional neural networks (DnCNNs) for image denoising [15]. In DnCNNs, the residual learning and batch normalization were used to speed up the training process and boost the denoising performance. DnCNNs is able to handle Gaussian denoising with unknown noise level. Authors in [16] proposed a very deep autoencoder for image restoration such as denoising and SR task. One of the main advantages of this work is the usage of skip connection to tackle the gradient vanishing problem. At video level, Lucas et al. in [17] presented a new generator network applied to the problem of video super-resolution.

One of the most recent attempts for quality enhancement techniques is the grand challenge at *Perceptual Image Restoration and Manipulation (PIRM)* workshop of ECCV 2018 [7] which aims to investigate models for SR task using different type of models and loss functions. An interesting discussion in this challenge is the difference between perceptual quality

and reconstruction accuracy. As discussed in [7], two trends are going on in the research linked to quality enhancement tasks. The first is to improve the reconstruction accuracy of an image using Full-Reference metrics. Using pixel-based metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) as loss function helps to reconstruct the image accurately, while it may result in low perceptual quality. Alternatively, the second trend is to improve the perceptual quality at the possible cost of lower reconstruction accuracy. During the past years, there has been a noticeable improvement in the reconstruction accuracy, either in terms of quantified metrics like Perception based Image Quality Evaluator (PIQE) [18] or perceptual ones (which are rated by users). With the increasing amount of advanced proposed SR methods, the disagreement between the two (reconstruction accuracy and perceptual quality) becomes more evident.

## III. DATASET AND EVALUATION METHOD

To build a dataset of lower quality frames for the training set, we extracted frames from encoded video(s) compressed using the H.264/AVC compression standard using the FFmpeg libx264 encoder wrapper. The videos were encoded in multiple resolution-bitrate pairs (bitrates ranging from 300 kbps to 2000 kbps and three different resolutions, 480p, 720p, and 1080p). For practical and quality estimation purpose, the 480p and 720p encoded video sequences were up-scaled to 1080p using the *bilinear* scaling function. In addition, the corresponding reference-quality frames are also extracted from uncompressed, high-quality reference videos and included in the training set. Hence, our training set consists of both high-quality reference frames as well as lower quality compressed and/or up-scaled frames. After frame extraction, we cropped $100k$ patches of size $96 \times 96 \times 3$ in RGB format from our training set. The initial selection of a smaller patch size (compared to increased patch sizes used in other similar works) was due to the hardware limitation, but as will be shown later in the discussion of the results, this already results in higher performance compared to other existing methods. The patches are cropped from 11k random frames extracted from multiple gaming videos. From each image, nine-patches are cropped randomly from nine regions of the image; from each region, one patch is selected. A uniform region pattern is chosen to cover a proper range of content, as shown in Figure 2.

For this work, we build three different datasets for three different research questions(in italics), hereinafter referred to as part-1, part-2 and part-3, to allow the reader to follow the paper better.

**Part-1**: The dataset is created based on 100k image patches which are extracted from a single game, League of Legends (LoL), but from multiple recorded video sequences representing various different levels and stages of the gameplay. This allows us to investigate *whether a game specific quality enhancement model can be built with high performance accuracy.*

Fig. 2: The selection of patches from one sample frame.

**Part-2**: The dataset is created based on 100k image patches which are extracted from 12 different video games, but also from multiple recorded video sequences of them as is done in Part-1. This is done to *investigate the potential of development of a generic quality enhancement model*, and its possible performance comparison with game-specific model (as is done in Part-1).

**Part-3**: The dataset is created based on 100k image patches which are extracted from one game, League of Legends, but consists of two sub-parts - first half consisting of patches from frames extracted from 480p and 720p videos up-scaled to 1080p videos; and the second half consisting of patches extracted from frames from 1080p videos encoded at various bitrate levels. This results in two sub-datasets with the first one corresponding to "blur" artefact and the second one consisting of "blockiness" artefact, which will allow us to compute the *trade-off between both these distortions* as is widely used in almost all adaptive streaming applications. In order to be fair in comparison, we first measured the quality of frames with blurriness using VMAF as a ground truth [19]. Next, we selected the corresponding frame with the closest VMAF value among those extracted for blockiness. The difference between the blockiness and bluriness was always ±10 VMAF values.

## IV. MODEL DEVELOPMENT

In this section, we introduce the proposed GAN architecture for the quality enhancement of frames from gaming videos. Our work uses the state-of-the-art GAN model, SRGAN [13], as the starting baseline model, which is then further modified to fit in with our objective with increased efficiency. We first describe the generative adversarial losses that are used in SRGAN. Then, the details on the SRGAN generator and discriminator networks are given, followed by information about our proposed generator networks.

### A. Loss Functions

The performance of the generator network relies strongly on the loss function. As discussed in [13], "pixel-wise loss functions such as MSE or PSNR (full-reference metrics) struggle to handle the uncertainty inherent in recovering lost

high-frequency details such as texture and minimizing MSE encourages finding pixel-wise averages of plausible solutions which are typically overly-smooth and thus have poor perceptual quality". Therefore, [13] proposed a new loss function, which is a weighted sum of the content loss and an adversarial loss. Content loss was defined to replace the pixel-based approach with a similar perceptual metric, which allows the model to recover better texture compared to the pixel-based models. For content loss, SRGAN uses the euclidean distance between the feature representations of a generated image and the reference image. Feature representations are the feature map of the last convolution before the max-pooling layer. The adversarial loss is then determined by the probability based on the discriminator over all training patches which allows us to differentiate between the generated patch and reference patch.

The main problem with content loss used in SRGAN is that VGG19 is trained on the ImageNET dataset, which is designed for a different task than the quality enhancement task. While usage of VGG19 was one of the significant innovation behind SRGAN, our preliminary results showed that retraining the VGG19 for quality enhancement task improves the results. Hence, we added another loss function to the model, named it distortion loss, which is based on retrained VGG19 network for quality prediction task. To retrain the VGG19 network for image quality prediction task, we use the frame-level VMAF values as our ground truth. The choice of VMAF is influenced by the fact that our earlier work has shown it to have a high correlation with subjective ratings for both gaming as well as non-gaming content [11]. Since the VGG19 network is initially designed for a classification task, the fully connected layer with multiple output neurons at the end of the network was removed in order to allow the model to get trained for the regression task. In addition, we added one dense layer consisting of only one output neuron with a linear activation. The output of the network was directly compared to the actual VMAF values of the validation set that includes different gaming videos, including those used in our training dataset. We cropped random patches of size $96 \times 96$ from the frames, in line with the patch size we use in the proposed GAN network. This was done in parallel to the training, such that in each epoch, a new random patch of each image was chosen. For training, we used transfer learning and freeze 75% of the VGG19 parameters, and only 25% were retrained. It must be noted that we do not use the output of the VGG19 network, which is the prediction of VMAF. Instead, we used the Euclidean distance between the feature representations of a generated image and the reference image. Our final loss function is a weighted average of SRGAN loss functions and retrained VGG19 loss (distortion loss).

### B. Network Architecture

SRGAN is designed for upscaling an image to a higher resolution without significant loss of quality. To allow SRGAN to be able to work for the enhancement of compression artifacts, we removed the pixel shuffler block, which upscales the image to a higher resolution. For the generator block, we

replaced the residual-based network in SRGAN with U-Net [20] network that can improve the flow of information as well as lead us to higher texture reconstruction. U-Net was used mostly for image segmentation of medical images, where the texture plays an important role, and there is limited input data available. We made a few changes for U-Net to improve the performance. We added the 'same' padding (zero paddings) instead of 'valid' padding (no padding). This change allows us to have the same input and output size, but using the 'same' padding introduces a minor ringing artifact near the borders. A possible solution to this could be mirror padding but is not explored in this paper and is left for future work. One of the advantages of using U-Net is having a skip connection, which allows the signal to be back-propagated to bottom layers directly and tackles the problem of gradient vanishing. Therefore, the skip connection allows deep training networks to achieve higher restoration performance. Another change compared to U-Net is to use different activation functions, where we utilize the Parametric Rectified Linear Unit (PReLU) function [21], instead of the ReLU or LeakyReLU as it has been shown that this activation function works better for compression related tasks [22].

## V. RESULTS

In the training process in contrast to the original U-Net, no data augmentation is used as we have enough input data in our training set. Due to the high number of input images, we used 100 epochs for training the data for each dataset. The best model was chosen based on the result of loss functions are stored. Next, we discuss first our (author) observations based on visual inspection of the obtained results of image enhancement, followed by results and observations based on an objective and subjective quality assessment study, supporting our initial observations.

### A. Observations

*1) Perceptual Quality vs. Image reconstruction:* As discussed before, there is a trade-off between the perceptual quality and image reconstruction accuracy, depending on the selection of the loss function used in the network. Our primary focus in this work was to improve the perceptual quality rather than the reconstruction accuracy. We observed that while, in general, the quality of the patches improves significantly, considerable improvement is observed in patches (frames) containing text, which became readable in the enhanced patches as compared to the non-readable text in the distorted patches. Using pixel-wise metrics such as PSNR or SSIM in the loss function allows the model to predict the text in the image well, while it fails to get high overall perceptual quality. For gaming content, while text, as well as numbers, form an important part of the video content, other parts such as the avatar/characters are equally (if not more) important, and hence quality enhancement on both aspects is of importance to our application. Therefore, we conclude that selection of loss function for video games is very game dependent.
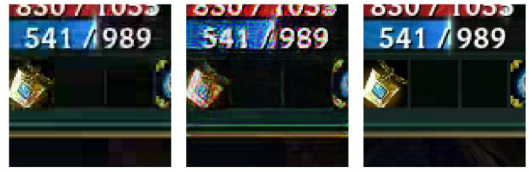


Fig. 3: Image enhancement in case of dealing with medium to high quality image. Left to right: Distorted patch, enhanced patch, reference patch.

*2) Add-up Distortion:* Another interesting observation is that the model tends to enhance the quality of a frame by generating fake data regardless of the quality level of a patch, which results in adding distortion to patches with medium to high-level quality values. This made us put a cut-off threshold, not to enhance the quality if the VGG19 loss is lower than a certain threshold. Figure 3 shows the additional distortion for a frame with high quality. As it can be seen the quality of enhanced image, is actually worse off than that of the distorted image.

### B. Objective and Subjective Measurement

In this section, we try to answer our research questions using objective and subjective quality assessment. To quantify the quality enhancement, we used NR quality metrics as it was proven that FR metrics are not a good means for perceptual quality measurement of enhancement techniques [7]. The main reason behind it is that a distorted image could be enhanced to a higher quality level without a high similarity to our reference image. As a choice of metric, we used PIQE [23] and Natural Image Quality Evaluator (NIQE) [24] which are frequently used as a measurement of quality enhancement of SR tasks [7].

*1) Enhancement Power:* We enhanced the distorted frames using our proposed model trained on dataset Part-1 to investigate the enhancement power. We selected 40 frames of dataset Part-1 in four classes of quality ranges as follows: Class 1 and 2 consisting of frames with Blockiness artifact with VMAF values range between 20 - 40 and 40-60 respectively, and Class 3 and 4 consisting of frames with Blurriness artifact with VMAF values range between 20 - 40 and 40-60 respectively.

Table I reports the quality in terms of NIQE and PIQE before and after enhancement of the distorted frames with different levels and types of distortion. It can be observed that the improvement for very low quality frames are more significant than the medium quality levels. While we had a distribution of distortion in the training set, we found that the enhancement would result in a certain level of quality, consequently, higher enhancement gained for low quality frames than medium quality frames. Figure 4 illustrates the before and after enhancement example frame from a gaming video using the proposed approach.

*2) Content Diversity:* Dataset Part-1 and Part-2 was used to investigate the impact of content diversity in training phase

Fig. 4: Before and after enhancement illustration using a single gaming video frame. The left half frame shows the video frame before enhancement while the second right half shows the frame quality after enhancement. Zoomed-in before and after enhancement patches of the image is shown on the right side for better clarity.

TABLE I: Quality comparison between distorted and enhanced frames of different quality/distortion classes in terms of NIQE and PIQE using the model developed for dataset part-1

|  | NIQE | | PIQE | |
| --- | --- | --- | --- | --- |
|  | Distorted | Enhanced | Distorted | Enhanced |
| Class-1 | 5.5 | 2.79 | 70.67 | 24.1 |
| Class-2 | 3.69 | 2.73 | 56.1 | 18.06 |
| Class-3 | 5.8 | 2.69 | 74.8 | 27.21 |
| Class-4 | 3.94 | 2.62 | 61.36 | 20.1 |

TABLE II: Quality comparison between distorted and enhanced frames of different quality/distortion classes in terms of NIQE and PIQE using the model developed for dataset part-2.

|  | NIQE | | PIQE | |
| --- | --- | --- | --- | --- |
|  | Distorted | Enhanced | Distorted | Enhanced |
| Class-1 | 5.5 | 3.21 | 70.67 | 38.23 |
| Class-2 | 3.69 | 3.32 | 56.1 | 36.41 |
| Class-3 | 5.8 | 3.11 | 74.8 | 34.25 |
| Class-4 | 3.94 | 3.91 | 61.36 | 30.22 |

TABLE III: Quality improvement between enhanced frames with regards to bluriness vs. blockiness in terms of NIQE and PIQE improvement using the model developed for dataset Part-1.

|  | NIQE Improvement | | PIQE Improvement | |
| --- | --- | --- | --- | --- |
|  | Bluriness | Blockiness | Bluriness | Blockiness |
| Class-1 | 2.68 | 2.21 | 45.15 | 37.36 |
| Class-2 | 1.54 | 1.14 | 34.35 | 28.16 |

on the performance of the model. Both datasets consist of 100k patches but one of them extracted from frames of only one game and the other extracted from 12 different games. We evaluate whether the model that is trained based on Part-1 (only LoL) performs better for new frames of LoL compared to the model trained for Part-2 or not. 40 test frames of LoL that are used the previous section was tested based on the two models. The result of enhancement using the model trained on Part-2, in terms of PIQE and NIQE is reported in Table II. Comparing Table I and II shows how important training dataset is for quality enhancement. On an average, approximately 12 PIQE value difference can be seen between game-specific model and general model (trained on multiple game). This is inline with findings of [12] that emphasizes on the importance of training set. However, it has to be noted that we are limited to the 100k patches used in the training set. Increasing the number of patches together with a deeper network might improve the results for the general model.

*3) Blurriness vs. Blockiness:* Dataset Part-3 was used to compare the performance of our enhancement model for blurriness and blockiness artifacts. Two models are built, one with images with blurriness, and one with the second part

consisting of only blockiness artifacts. In order to test the models, we selected 40 pairs of LoL frames with the same quality level in terms of VMAF, one with blockiness and the other with blurriness artifacts. The frames are classified into two classes depending on their quality range, class-1 with ranging quality from 20 to 40 VMAF values, and class-2 with VMAF value ranges from 40 to 60. The result in terms of objective quality metric is reported in Table III. As it can be seen, frames with blurriness artifacts can be improved better than blockiness which might be due to uniform artifacts.

*4) Subjective Quality Assessment:* In order to quantify the perceptual enhancement of our model, we conducted a short
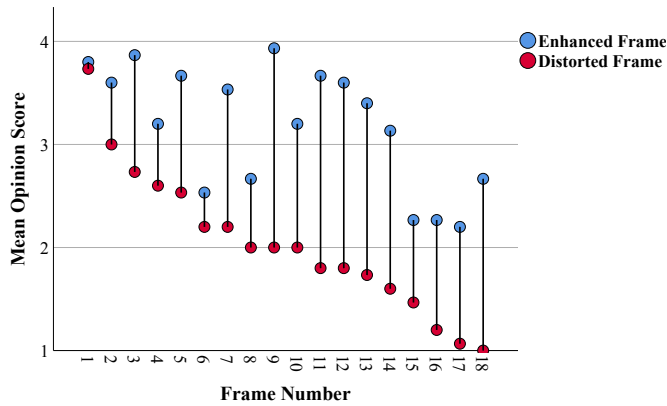
Fig. 5: Mean Opinion Score difference between enhanced and distorted frames.

subjective test with 15 subjects. We selected 2 random frames from the game LoL and used the model trained on dataset part-1 to test our model. Eight distorted frames are selected from each reference frames which have different types and levels of distortion. In total, 38 images including 18 distorted images, 18 enhanced images and 2 reference images are used in a the subjective test adhering to ITU-T Recommendation P.910 [25]. For the subjective test, in addition to the image quality rating, we asked participants to rate the level of blurriness and blockiness in the image in 5-point ACR scale. This gives us better understanding about the distortion after the enhancement. PIQE and NIQE scores results in -0.77 and -0.64 Pearson Correlation coefficient (PCC) score respectively with respect to the mean opinion score (MOS). Figure 5 plots the difference of MOS between distorted and corresponding enhanced frames in order to show the perceptual enhancement power of the proposed method. From Figure 5 we can observe a maximum 2 MOS difference (in frame 9, affected by blur). The enhancement in terms of MOS for frames with blockiness (frame 1, 2, 4, 6, 8, 10, 15, 16, 17) is lower than that observed for the frames with blurriness artifact (rest of the frames). The reason behind this can be attributed to the fact that blurriness is more uniformly distributed and hence easier to learn and also to the fact that the baseline SRGAN model is designed to minimise such artifact.

## VI. DISCUSSION AND CONCLUSION

In this paper, we investigated the performance of deep learning methods for enhancement of gaming content. We proposed a new model based on state-of-the-art models for super-resolution task and investigated the importance of distortion and training dataset to build a high performance model. This is the first work, to the best of authors knowledge, where the enhancement is applied on gaming frames. The result of this work can help guide the research community to design game-specific enhancement models. It has to be noted that the current work was limited to frame level enhancement and we leave the application to videos level as a future work.

REFERENCES

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2017–2022," https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html, Feb 2019.

[2] Netflix, "Netflix Per-Title Encode Optimization," https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2, Dec 2015, [Online: accessed 15-Jan-2020].

[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[6] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[7] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 pirm challenge on perceptual image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.

[8] Z. Zhao, Q. Sun, H. Yang, H. Qiao, Z. Wang, and D. O. Wu, "Compression artifacts reduction by improved generative adversarial networks," *EURASIP Journal on Image and Video Processing*, vol. 2019, no. 1, pp. 1–7, 2019.

[9] N. Barman and M. G. Martini, "H.264/MPEG-AVC, H.265/MPEG-HEVC and VP9 codec comparison for live gaming video streaming," in *Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, Erfurt, Germany, May 2017, pp. 1–6.

[10] N. Barman, S. Zadtootaghaj, M. G. Martini, S. Möller, and S. Lee, "A Comparative Quality Assessment Study for Gaming and Non-Gaming Videos," in *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, Sardinia, Italy, May 2018.

[11] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Möller, "An Evaluation of Video Quality Assessment Metrics for Passive Gaming Video Streaming," in *Proceedings of the 23rd Packet Video Workshop*, ser. PV '18, Amsterdam, Netherlands, 2018, pp. 7–12.

[12] N. Takano and G. Alaghband, "Srgan: Training dataset matters," *arXiv preprint arXiv:1903.09922*, 2019.

[13] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.

[14] C. Dong, Y. Deng, C. Change Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 576–584.

[15] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[16] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Advances in neural information processing systems*, 2016, pp. 2802–2810.

[17] A. Lucas, A. K. Katsaggelos, S. Lopez-Tapuia, and R. Molina, "Generative adversarial networks and perceptual losses for video super-resolution," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, Oct 2018, pp. 51–55.

[18] N. Venkatanath, D. Praneeth, M. C. Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC)*. IEEE, 2015, pp. 1–6.

[19] Netflix, "VMAF - Video Multi-Method Assessment Fusion," https://github.com/Netflix/vmaf, [Online: Accessed 2-Jan-2020].

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[22] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, "Deep convolutional autoencoder-based lossy image compression," in *2018 Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 253–257.

[23] Venkatanath N, Praneeth D, Maruthi Chandrasekhar Bh, S. S. Channappayya, and S. S. Medasani, "Blind image quality evaluation using perception based features," in *2015 Twenty First National Conference on Communications (NCC)*, Feb 2015, pp. 1–6.

[24] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "Completely Blind" Image Quality Analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, March 2013.

[25] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommunication Union, Geneva, 2008.