QUALITY OF VIDEO RENDERING TECHNIQUES USING ARTIFICIAL INTELLIGENCE

D.K. Mohanty[1], G.R. Thippeswamy[2], G. Erappa[3] and Vishal Gangadhar Puranik[4]
[1]*Government B.Ed. Training College Kalinga, India*
[2]*Department of Computer Science and Engineering, Don Bosco Institute of Technology, India*
[3]*Department of Information Science and Engineering, RR Institute of Technology, India*
[4]*Department of Electronics and Telecommunication Engineering, JSPM's Bhivarabai Sawant Institute of Technology and Research, India*

Abstract

In this paper, we propose a novel method that makes use of artificial intelligence to determine in a quick and accurate manner which bitrate ladder is best suited to each specific video scenario. This method is included as part of our overall contribution to this body of research. To accomplish fast entropy-based scene recognition using the artificial intelligence technique, a CNN model is utilised as part of the overall strategy. We were able to significantly reduce the amount of processing time necessary to recognise the scenes because we were dealing with versions of the video sequences that had both a lower quality and a lower bitrate. This allowed us to work more quickly. We first generated a training dataset that was large enough to train a convolutional neural network utilising the x264 video codec, and then we used that dataset to generate multiple encodings with varying bitrates, presets, and resolutions. The training dataset was created using the x264 video codec. As a result of the research that we carried out, we concluded that a particular collection of input features for the CNN model can be used to acquire a more accurate prediction of the level of video quality that will be produced. By predicting the PSNR quality measure for the segments, the suggested CNN model brings down the MAE and MSE to 0.2 and 0.05, respectively. This is accomplished by reducing the number of segments. This serves to reduce the amount of error overall.

Keywords:

Artificial Intelligence, Convolutional Neural Network, Video Quality Enhancement, Video Rendering

1. INTRODUCTION

There is an ecosystem of individualised video streaming applications and services that is continuously growing in scope. This ecosystem is referred to as the internet of things. The bar that must be cleared to achieve this objective is raised by the capability of encoding, disseminating, sharing, and consuming video streams on any device and at any moment. Streaming platforms make it their mission to deliver the best possible experience to users while they are streaming by considering not only the preferences of the user but also the quantity of bandwidth that is currently available [1].

To accomplish this objective, suppliers and services of streaming media utilise a wide range of cutting-edge technologies in addition to industry standards. These technologies and standards include things like HTTP Adaptive Streaming (HAS) [2], video codecs [3], and optimised bitrate ladder calculations for video file formats, to name a few instances each. When using HAS, a video stream is split up into numerous representations, and then each of those representations is sent to a client device in increments that range from a few seconds to a few minutes.

These transmission intervals are determined by several variables, some of which are, for example, the available bandwidth and the resolution of the display. Because of advancements in HAS technology, it is now feasible to switch between accessible representations to make up for variations in the performance of the network. As demonstrated by the MPEG-DASH demonstration, HAS can been implemented without restricting the user choice of video format at any point in the process [5].

Use of a concept known as one size fits all (also known as a fixed/static/classic bitrate ladder) is the strategy to selecting bitrate-resolution pairs that is known to be the most straightforward and all-encompassing approach. A static bitrate ladder has representations that are predetermined and fixed. These representations encompass a variety of different possible levels of visual quality by pairing different bitrates with different resolutions.

This form of bitrate ladder is also known as a fixed bitrate ladder in some circles. When the complete video is encoded with a single bitrate and resolution pair, however, the visual quality of the encoded portions is different [6]. The quality of the original video is diminished during the encoding procedure, which explains why this occurs. The reality that the bitrate has not changed impedes the video encoder ability to produce higher quality output, which in turn prevents the value from increasing. This is since the complexity of the video image is continually increasing.

To determine the optimal bitrate ladder for a video sequence, utilise video complexity analysis, test (or trial) encodings, and machine learning techniques. This is done to ensure that the video sequence receives the optimum bitrate ladder when it is compressed. Per-title encoding was Netflix first attempt at optimising video quality measures such as PSNR and VMAF This was accomplished by selecting optimal bitrates for each individual video based on the degree of difficulty that it presented. The per-title bitrate ladder is superior to the constant bitrate ladder in terms of both the quality of the general product and the amount of space it saves [7].

The per-title encoding scheme is also known as the per-shot, per-scene, and context-aware encoding schemes. These variants include When using these techniques, the bitrate-resolution hierarchy for a video must still be determined by encoding the video with many different resolutions and bitrates. Calculations generally take longer and cost more money because of the increased number of processing units that are used, the vast majority of which are stored in the cloud. This is because this increased number of processing units is required.

Additionally, most approaches determine the bitrate ladder for the duration of the video. This is done even though a single video may comprise multiple video scenes and even segments that have a significantly different level of the overall level of visual complexity [8]. Because of this, it is recommended that the movie be divided into sequences that contain a comparable amount of visual complexity so that each can be evaluated on its own. to reiterate, computational tools and an in-depth frame-by-frame analysis of a high-definition video are necessary to divide the video into scenes that have varying degrees of difficulty.

We suggest a novel and accurate method for reducing the amount of time spent detecting scenes and calculating an optimised bitrate ladder. This method should result in a time savings. This technique, which consists of two stages and is known as entropy-based scene detection and machine learning, has been given its name. Entropy-based scene identification and scene-specific bitrate ladder optimization are the two primary foci of this research. 1) Entropy-based scene identification and 2) scene-specific bitrate ladder optimization to reduce the amount of time spent on the computation process, an artificial intelligence builds a temporal information metric and its associated entropy for video sequences that are encoded using a low bitrate and resolution. Our strategy is predicated on the findings of prior research that have demonstrated a significant connection between the temporal information of the original movies and the low bitrate and resolution encodings that were used for those movies. These findings have served as the foundation for our approach. When it is feasible for us to do so, we use artificial intelligence to optimise the encoding for each unique circumstance and to provide adaptive streaming video with descriptions of a wide variety of media presentations [9].

In recent years, one of the scientific approaches that has been use in the real world most rapidly has been artificial intelligence, also known as AI. This has been the case for several different scientific approaches. This capability is made possible by the techniques of artificial intelligence. High-performance computing and increased data storage capacities have enabled and accelerated the widespread adoption of artificial intelligence technologies in a wide variety of fields, ranging from the mundane to the highly specialised, such as finance and the financial sector, as well as national security and command and control. These fields include finance and the financial industry. These developments have enabled and accelerated the widespread adoption of AI technologies in a broad variety of fields, which opens a lot of new possibilities.

2. BACKGROUND

The ability of computers and other electronic devices to carry out activities that would typically require the intelligence of humans is what is meant when we talk about artificial intelligence. These tasks include reading aloud and comprehending written text, recognising spoken language, and providing a reaction to it, recognising images and determining what they depict, and even predicting what will happen in the future. Artificial intelligence has been used on a more advanced level to research human behaviour by overhearing conversations and observing how people interact with one another in social situations. In addition to its use in the forecasting of natural occurrences, it has also been used to gain an understanding of socially significant problems such as homelessness [10].

Governments in every part of the world have come to the realisation that artificial intelligence has the potential to hasten not only the rate of economic development but also the rate of social advancement. Nevertheless, to fully realise the potential of artificial intelligence technologies, legitimate societal concerns need to be taken into consideration during the process of developing and implementing these technologies. This is necessary to fully realise the potential of AI technologies.

When compared to the processes that are associated with enterprises that are more mundane, those that are associated with creativity require a very different kind of originality and expertise. In contrast to the success of artificial intelligence, which is dependent on data conformity, creativity is propelled by human imagination and may defy general principles. Because of this capacity, they can consider possibilities that they were previously unable to even begin considering. cannot be easily fixed by the limited machine learning algorithms that are accessible now.

Researchers have dedicated a significant amount of time and effort over the course of several decades to investigating the possibility that intelligence could play a role in the creative process. In the past, there was a widespread misunderstanding that artificial intelligence would endeavour to imitate human ingenuity. This led to a barrier for the development of artificial intelligence. This exemplifies a general awareness of the current state of the art and demonstrates a widespread understanding of artificial intelligence as a tool. Most of the AI technologies have been designed to operate within closed domains, where they can assist and support people rather than replacing them. Because of this, the advantages of the synergy can be maximised through increased collaboration between human and artificial intelligence technologies.

Artificial intelligence is being put to greater use across a variety of business sectors, including advertising and marketing, the gaming and immersive application industries, and advertising and marketing, according to a study of business usage performed by Crunchbase. They talk about the present state of research and development in the field of artificial intelligence and machine learning, as well as forthcoming challenges and trends in the industry.

3. AI BASED VIDEO RENDERING

It is becoming more standard practise to employ convolutional networks, which are a subclass of deep feed forward CNNs. They are built from a succession of convolutional layers that are optimised for utilising two-dimensional structures, such as those that are found in images. These structures can be found in computer models which is shown in Fig.1.

These make use of convolutional processes, and the output of each convolutional layer is the incoming signal after it has been put through a convolution filter and transformed in some way. The kernel that is utilised in these processes has a predetermined size, and the internal signal that is utilised in these procedures also has a predetermined size. During the training process, the weights of the filter are modified in response to a loss function. This function assesses the degree to which the network projections deviate from the actual world. This assessment is carried out by contrasting the predictions made by the network with the data that was collected.

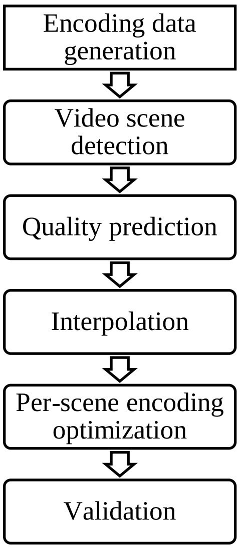

Fig.1. Proposed Framework

Several strategies, such as ℓ_1, ℓ_2, SSIM, and perception loss, are utilised on a consistent basis. to fine-tune the filter weights and backpropagate the errors, the estimated gradients of the local error surface are used in a series of consecutive forward and backward cycles. The determines which features are picked up, linking the features that have been recognised to the characteristics that were present in the training data. In other words, this establishes which features are picked up. The initial stages of CNNs oversee the process of extracting the most fundamental visual information. This method is comparable to the visual basis processes that are in the primary visual cortex of the brain.

A typical CNN will have the configuration shown in Figure 3, which connects the outputs of the convolution layer to a pooling layer. This is the most prevalent configuration for a CNN. This layer combines the information obtained from several different neuron clusters into a singular neuron. Following this step, activation functions like tanh and ReLU are applied to the linear network to transform it into a non-linear structure. It is possible that successive iterations of this pattern will maintain the same kernel dimensions, but it is also possible that they will change.

During the first layer of processing, the CNN acquires the knowledge necessary to recognise edges in the raw images. It then applies this knowledge during the second layer of processing to recognise fundamental forms. The higher-level characteristics that are generated by the more complex layers have substantially increased semantic significance as a result. When it comes to the actual implementation, the classification function of the network is not carried out until the very final few levels. These layers consist of a fully connected one and a softmax one, which scales the output in an exponential fashion between 0 and 1, representing the probability distribution that corresponds to the class that is supposed to be predicted.

One of the most popular and extensively used backbone networks right now is called VGG, which offers users a choice between two distinct variants (16 and 19 layers, respectively). The networks are built with a sequence of convolution blocks in a sequential order (which includes convolutional layers, ReLU activations, and a max-pooling layer), and the ReLU activations are used to connect the last three layers of the network altogether.

The designs of VGG neural networks are more complicated than those of earlier generations of neural networks because they make use of a reduced number of receptive fields than those of earlier generations (3×3, with a stride of 1). The VGG network that is used by DeepArt has levels that are not completely connected to one another, and this is done on purpose. It is standard procedure to incorporate pre-trained VGG networks into the design processes of other networks to conduct perception loss analysis (and style loss).

3.1 ENCODING

The selection process at this stage involves choosing videos with varying degrees of complexity; these videos will have sequences of video frames that contain various types of information regarding both space and time. You can determine a movie spatial information, which is a metric that measures the degree of spatial complexity, by using the following equation.

$$SI = \max \forall F_n\{\sigma[\text{Sobel}(F_n)]\} \tag{1}$$

where

F_n - video frame

n - time, and

σ - standard deviation

When it comes to video communications, the quantification of luminosity, which is also referred to as luminance, is the responsibility of the luminance component. After that, the max function is used to find the frame in the complete movie that has the highest standard deviation by comparing it to the others.

The temporal information (TI) of a video represents the quantity of motion that is present in the video. This information is determined by applying a motion difference function Mn between the luminance components of identically spaced pixels in two sequential frames F_n and F_{n-1}:

$$M_n(i,j) = F_n(i,j) - F_{n-1}(i,j), \tag{2}$$

where

$F_n(i,j)$ - frame pixel

i - row and

j - column

n - time.

The TI measure is composed of the highest standard deviation of $M_n(i,j)$ that is computed across all the pixels. This variance is taken across the entire image.

$$TI = \max\{\sigma[M_n(i,j)]\}. \tag{3}$$

When adaptive video streaming is utilised, the duration of each video segment winds up becoming an extremely essential quality since adaptive and dynamic switching between representations typically begins at segment boundaries.

The subsequent stage is to snip the chosen video sequences into bite-sized pieces that are anywhere from two to four seconds long. Both in the business world and in the scholarly world, this is the standard measurement for bite-sized pieces. When compared to the 2s segments, which are generally used for low-latency streaming, the 4s segments demonstrate an acceptable trade-off between encoding efficiency and video streaming performance.

This is because the 4s segments have more information packed into each bit of data. This can be seen by looking at how frequently the 2s portions are used, which is a demonstration of the point. The encoded video chunks that make up the dataset are then dissected to locate the characteristics that are of the greatest significance to extract useful information from them. These might include the labels of the segments, the input sizes, the resolutions, and so on.

3.2 QUALITY PREDICTION

The research focus on the measure of video segments that is known as the Luminance Peak Signal-to-Noise Ratio (LPSNR). We base our estimations of the video overall quality on the YPSNR score, which is a metric that is universally acknowledged as being objective in its measurement of video quality. This score is derived from the ratio of the peak signal to noise ratio (PSNR) to the average SNR.

The dataset that was generated in the first portion is subjected to some preliminary processing before we begin training the CNN. Finding the encoding preset those results in the lowest YPSNR requires testing each conceivable combination of segment name, width, and height, as well as the encoding preset. This allows us to cut down on the number of entries contained in the dataset. Following that, the data were partitioned into a training set and a test group to conduct further research on them.

During the process of developing the CNN, we concluded that a sequential approach would be most appropriate. Only the first input layer in this model is given the actual data, while the concealed levels oversee automatically recognising input patterns based on the data that is given to them. Then, as we were putting it through its training, we decided upon the most efficient trigger mechanisms to apply.

We used a linear activation function for the final layer (the one that the viewer sees), but for the intermediate and concealed levels, we used a non-linear activation function known as the rectified linear unit, or ReLU.

The ReLU function will carry out a half-rectification whenever it is presented with a negative input, which will result in the integer being set to 0. Both the mean squared error (MSE) and the MAE, which are both presented as averages of the disparity between the observed and predicted values, respectively, were computed by us. These computations were performed with the intention of determining how efficient CNN output was.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_i - \bar{y}_i\right| \tag{4}$$

$$MSE = \frac{1}{n}\sum_{j=1}^{n}\left(y_i - \bar{y}_i\right)^2 \tag{5}$$

where,

n - predicted qualities,

y_j - predicted video quality and

y'_j - actual video quality.

By iteratively tuning MAE and MSE and updating the training features of the CNN, we can accomplish high levels of reliability and precision. In the end, we make a projection about the YPSNR across all the dataset inputs by employing the CNN model that we constructed and training it with the data which are mentioned in Table 1 to Table 7.

Table.1. MAE

Videos	CNN Quality Prediction	VGG-16 Quality Prediction	VGG-19 Quality Prediction
Video 1	4.520	4.510	4.485
Video 2	4.500	4.505	4.495
Video 3	4.326	4.414	4.343
Video 4	3.242	3.146	3.242
Video 5	3.283	3.263	3.460
Video 6	3.106	3.187	3.379
Video 7	3.343	3.359	3.535
Video 8	3.677	3.742	3.652
Video 9	3.742	3.611	3.495
Video 10	3.434	3.480	3.586
Video 11	3.732	3.859	3.904
Video 12	3.949	3.798	3.949
Video 13	3.788	3.601	3.904
Video 14	3.808	3.596	3.803
Video 15	4.136	4.126	3.859
Video 16	3.237	3.338	2.939
Video 17	3.455	2.768	3.374
Video 18	3.985	3.460	3.556
Video 19	4.025	4.056	4.131
Video 20	3.444	3.197	3.419
Video 21	3.520	3.601	3.576

Table.2. MSE

Videos	CNN Quality Prediction	VGG-16 Quality Prediction	VGG-19 Quality Prediction
Video 1	4.395	4.434	4.445
Video 2	4.439	4.454	4.453
Video 3	4.235	4.295	4.283
Video 4	3.259	3.373	3.241
Video 5	3.239	3.528	3.224
Video 6	3.219	3.064	3.166
Video 7	3.104	3.353	3.258
Video 8	3.627	3.383	3.531
Video 9	3.438	3.563	3.568
Video 10	3.518	3.398	3.439
Video 11	3.911	3.807	3.780
Video 12	3.777	3.881	3.797
Video 13	3.762	3.288	3.594

Video 14	3.727	3.817	3.714
Video 15	4.121	3.229	3.888
Video 16	3.438	3.448	3.125
Video 17	3.284	3.732	3.319
Video 18	3.284	3.652	3.646
Video 19	3.891	4.046	3.989
Video 20	3.209	3.184	3.329
Video 21	3.533	3.642	3.536

Table.3. PSNR

Videos	**CNN Quality Prediction**	**VGG-16 Quality Prediction**	**VGG-19 Quality Prediction**
Video 1	49.94	49.89	49.56
Video 2	49.83	49.89	49.89
Video 3	48.69	47.24	47.10
Video 4	35.06	37.39	36.78
Video 5	34.39	34.39	33.33
Video 6	33.78	38.67	33.56
Video 7	33.56	36.33	36.11
Video 8	40.61	36.78	38.06
Video 9	40.06	42.17	38.61
Video 10	37.89	39.11	37.22
Video 11	41.39	42.44	40.89
Video 12	40.72	42.11	41.78
Video 13	39.17	42.22	36.39
Video 14	38.28	43.06	42.56
Video 15	45.89	41.56	44.11
Video 16	34.67	31.89	30.78
Video 17	39.94	37.72	34.67
Video 18	39.78	43.67	44.00
Video 19	44.78	43.56	45.06
Video 20	39.22	36.89	39.33
Video 21	40.67	36.83	40.22

Table.4. Accuracy

Videos	**CNN Quality Prediction**	**VGG-16 Quality Prediction**	**VGG-19 Quality Prediction**
Video 1	94.395	94.290	93.660
Video 2	94.185	94.290	94.290
Video 3	92.022	89.292	89.019
Video 4	66.255	70.665	69.510
Video 5	64.995	64.995	63.000
Video 6	63.840	73.080	63.420
Video 7	63.420	68.670	68.250
Video 8	76.755	69.510	71.925
Video 9	75.705	79.695	72.975
Video 10	71.610	73.920	70.350
Video 11	78.225	80.220	77.280
Video 12	76.965	79.590	78.960
Video 13	74.025	79.800	68.775
Video 14	72.345	81.375	80.430
Video 15	86.730	78.540	83.370
Video 16	65.520	60.270	58.170
Video 17	75.495	71.295	65.520
Video 18	75.180	82.530	83.160
Video 19	84.630	82.320	85.155
Video 20	74.130	69.720	74.340
Video 21	76.860	69.615	76.020

Table.5. Precision

Videos	**CNN Quality Prediction**	**VGG-16 Quality Prediction**	**VGG-19 Quality Prediction**
Video 1	93.975	93.765	93.240
Video 2	93.555	93.660	93.450
Video 3	89.933	91.770	90.300
Video 4	67.410	65.415	67.410
Video 5	68.250	67.830	71.925
Video 6	64.575	66.255	70.245
Video 7	69.510	69.825	73.500
Video 8	76.440	77.805	75.915
Video 9	77.805	75.075	72.660
Video 10	71.400	72.345	74.550
Video 11	77.595	80.220	81.165
Video 12	82.110	78.960	82.110
Video 13	78.750	74.865	81.165
Video 14	79.170	74.760	79.065
Video 15	85.995	85.785	80.220
Video 16	67.305	69.405	61.110
Video 17	71.820	57.540	70.140
Video 18	82.845	71.925	73.920
Video 19	83.685	84.315	85.890
Video 20	71.610	66.465	71.085
Video 21	73.185	74.865	74.340

Table.6. Recall

Videos	**CNN Quality Prediction**	**VGG-16 Quality Prediction**	**VGG-19 Quality Prediction**
Video 1	92.610	93.450	93.671
Video 2	93.555	93.870	93.849
Video 3	89.250	90.510	90.258

Video 4	68.670	71.085	68.303
Video 5	68.250	74.340	67.946
Video 6	67.830	64.575	66.728
Video 7	65.415	70.665	68.649
Video 8	76.440	71.295	74.403
Video 9	72.450	75.075	75.180
Video 10	74.130	71.610	72.482
Video 11	82.425	80.220	79.664
Video 12	79.590	81.795	80.010
Video 13	79.275	69.300	75.737
Video 14	78.540	80.430	78.257
Video 15	86.835	68.040	81.932
Video 16	72.450	72.660	65.856
Video 17	69.195	78.645	69.951
Video 18	69.195	76.965	76.839
Video 19	82.005	85.260	84.053
Video 20	67.620	67.095	70.161
Video 21	74.445	76.755	74.508

Table.7. F-Measure

Videos	**CNN Quality Prediction**	**VGG-16 Quality Prediction**	**VGG-19 Quality Prediction**
Video 1	98.168	97.505	98.390
Video 2	98.390	98.500	98.832
Video 3	95.073	93.968	95.294
Video 4	70.973	72.300	74.842
Video 5	75.727	71.858	78.269
Video 6	73.958	71.415	67.988
Video 7	77.385	68.873	74.400
Video 8	79.928	80.480	75.063
Video 9	76.501	76.280	79.043
Video 10	78.491	78.048	75.395
Video 11	81.365	81.696	84.460
Video 12	83.134	86.450	83.134
Video 13	72.410	82.913	78.822
Video 14	84.681	83.355	78.712
Video 15	87.777	90.540	90.319
Video 16	61.245	70.863	73.074
Video 17	68.983	75.616	60.581
Video 18	87.556	87.224	75.727
Video 19	89.656	88.108	88.772
Video 20	78.269	75.395	69.978
Video 21	80.038	77.053	78.822

We make use of both predicted and interpolated convex hulls at this stage of the process to establish which bitrate ladders produce the best results for each individual video recording. If you select all the points on the interpolated convex hull that are located within the acceptable range of YPSNR values, you will be able to achieve the highest resolution possible for each individual video scene. We make use of the video quality spacing points that were decided upon earlier to arrive at an accurate determination of the number of bitrate resolution pairs and values of quality objectives that should be used for each scene. This allows us to arrive at an accurate determination of the number of bitrates and resolutions that should be used for each scene. Using an interpolated convex hull, the software determines the bitrates and resolutions that will result in the highest possible quality for a given collection of quality points (or targets).

4. PERFORMANCE EVALUATION

The TID 2008, TID 2013, and LIVE databases were utilised in the fourth step, which was an analysis of the successes achieved through training and testing. Every one of the suggestions made for a solution was put through its tests (Table.1-Table.7). For the purposes of IQA research, it is an acceptable and prevalent practise to make use of image collections such as these.

There is a mean opinion number as well as information about the sort of distortion and the degree of distortion associated with each image. This evaluation is a weighted estimate based on the viewpoints expressed by a variety of individuals. Although every image in the TID databases has a 512x384, the aspect ratios of the images in the LIVE collection are all different.

In this segment, we evaluate our per-scene compression technique by comparing it to the traditional bitrate ladder and judging it based on three criteria: file size, transfer speed, and quality. By comparing the bitrate numbers that are utilised at the various stages of the bitrate ladder, we can compute the bitrate reduction Br metric. This allows us to determine how much the bitrate has decreased.

To be more particular, we used a scene detection algorithm that was suggested to us to recognise and divide up the various scenes that were contained within video clips. Our technique recognises the beginning of a new scene when the average and entropy of the TI metric for each frame of the video sequence reach a threshold that has been established in advance. This threshold was determined before the video sequence was analysed. The TI values that are a part of our raw encoding information are utilised by the algorithm that is responsible for scene identification. These numbers apply to movies that are encoded at a frame rate of 144 frames per second.

Because the Keras Python library allows for independent implementation, we were able to independently implement the CNN model for each of the input groups. Because of this, we were able to make some educated guesses about the YPSNR quality of the middle sections of the video sequences that were found. Every distinct CNN model has between seven and eleven input neurons, seven hidden levels, and just one output neuron. The performance of the Adadelta optimization algorithm was compared to that of three other optimization algorithms, namely Adagrad, Adadelta, and RMsprop.

Based on the findings of this comparison, the Adadelta optimization algorithm was selected as the best option for use in the construction of the CNN model. An enhanced variant of the Adagrad optimizer, the Adadelta optimizer was developed by Google. It optimises the learning rate hyperparameter based not on a complete history of gradients but rather on a rolling window of gradient changes. This contrasts with other optimization methods, which base their decisions on the entire history of gradients. Even when it is subjected to numerous updates, other optimization algorithms are unable to match Adadelta ability to learn and adjust the default values of its learning parameters. This ability distinguishes Adadelta from other optimization algorithms.

It is possible to ascertain the differences that exist between the many different individual frames by utilising the TransNet model. You will be able to characterise the nature of changes, such as whether they are sudden or gradual, because of this ability. As input, the model will take into consideration a series of images that each have a pixel depth of 24 bits. If you were wondering, the granularity of input layer networks is 48x27, but just in case you weren't: This makes it feasible to load all the movies into RAM in many situations, which enables the processing to take place more quickly. The probabilities can be utilised in this manner to generate a collection of shot limits, which can then be input into any evaluative algorithm. It is possible to use these shot limits to ascertain whether an event has taken place.

It conceivable that the problem lies with the implementation-specific refining techniques that were used in this investigation; that something that needs to be investigated further. We altered the model input shape to make calculations run more quickly on graphics processing units (GPUs). It is possible that because of this, the aspect ratio of some of the images will change in an unusual fashion. This could happen if the images are resized. Despite this, the primary objective of this investigation is to arrive at a more accurate proportional number. The basic accuracy of TID2008 was improved by approximately 20% because of an additional pre-training step. On the other hand, it seems as though the F1 score as well as the precision of the classification are on the lower part of the spectrum.

5. CONCLUSION

The pre-training phase is maintained, even though the classifier model was discarded due to the potential impact that it could have on the IQA result. The reason for this was because the pre-training phase could be used to train the classifier model. The content-based patching assessment strategy functions more effectively than the baseline approach for the TID2013 and LIVE datasets. This difference is due to the content-based patching assessment strategy focus on improving the quality of the content. An intriguing aspect of the research is the fact that the content-based patching had such a minimal effect on the TID 2008 baseline correlation ratings. It possible that this is a sign that there something wrong with the way the detection is working. This underwent manual testing, which revealed that the accuracy of the proposed framework is affected by a broad variety of factors and parameters.

REFERENCES

[1] Giuseppe Baruffa and Fabrizio Frescura, "Adaptive Error Protection Coding for Wireless Transmission of Motion JPEG 2000 Video", *EURASIP Journal on Image and Video Processing*, Vol. 10, pp. 123-134, 2016.

[2] S. Ponlatha and R.S. Sabeenian, "Comparison of Video Compression Standards", *International Journal of Computer and Electrical Engineering*, Vol. 5, No. 6, pp. 549-554, 2013

[3] M. Deepa and M.C. Binish, "A Fast Intra Prediction for H.264/AVC based on SATD and Prediction Direction", *Proceedings of International Conference on Emerging Trends in Engineering, Science and Technology*, pp. 1016-1023, 2016.

[4] V. Bichu, G. Hegde and S. Sanju, "Fast Block-Matching Motion Estimation using Modified Diamond Search Algorithm", *Proceedings of International Journal of Advanced Computer Engineering and Communication Technology*, pp. 423-429, 2014.

[5] M. Ma, O. C. Au and S.H.G. Chan, "Edge-Directed Error Concealment", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 20, No. 3, pp. 382-395, 2010.

[6] Hadi Asheri, Hamid R. Rabiee, Nima Pourdamghani and Mohammad Ghanbari, "Multi-Directional Spatial Error Concealment using Adaptive Edge Thresholding", *IEEE Transactions on Consumer Electronics*, Vol. 58, No. 3, pp. 880-885, 2012.

[7] Ulil S. Zulpratita, "GOP Length Effect Analysis on H.264/AVC Video Streaming Transmission Quality over LTE Network", *Proceedings of International Conference on Computer Science and Information Technology*, pp. 5-9, 2013.

[8] K. Asha, D. Anuradha and M. Rizvana, "Human Vision System's Region of Interest Based Video Coding", *Compusoft*, Vol. 2, No. 5, pp. 127-134, 2013.

[9] C. Chandrasekar, "Qos-Continuous Live Media Streaming in Mobile Environment using VBR and Edge Network", *International Journal of Computer Applications*, Vol. 53, No. 6, pp. 1-13, 2012.

[10] Bruno Zatt, Marcelo Schiavon Porto, Jacob Scharcanski and Sergio Bampi, "GOP Structure Adaptive to the Video Content for Efficient H.264/AVC Encoding", *Proceedings of International Conference on Image Processing*, pp. 281-288, 2014.