Al-Noor Journal of Engineering Management and Computer Science
ISSN: 3079-0689 (Online)
https://njemcs.edu.iq/index.php/njemcs/

Video Compression and Redundancy Reduction by Hybrid Neural Networks (CNN+RNN(LSTM)) Algorithm and Generative Adversarial Network (GAN) Algorithm

Sama Jassim Mohammed [1], Waleed Abdullah Araheemah [2]

[1,2] Technical College of Administration - Baghdad Middle Technical University /Baghdad, Iraq.

ARTICLE INFO

Article history:
Received 24 July 2024
Revised 24 July 2025
Accepted 10 August 2025
Available online 01 September 2025

Keywords:
Video Compressive
Convolutional Neural Networks (CNN)
Recurrent Neural Networks (RNN)
Long Short-Term Memory (LSTM)
compression ratio
Peak signal-to-noise ratio (PSNR)

ABSTRACT

As the need for high-quality video increases it has become necessary to develop smarter compression methods. Artificial neural networks (ANNs) have emerged as an effective tool in this field enabling them to extract the most important features from video and minimize temporal and spatial redundancy resulting in an efficient and high-quality compressed representation. Some of the most prominent models used are: Hybrid neural networks (CNN+RNN(LSTM)) Hybrid neural networks combine convolutional neural networks (CNNs), recurrent neural networks (RNNs), and especially long short-term memory (LSTM) units to extract spatial and temporal features from data, GAN, GANs compress video files while maintaining high perceptual quality. They encode each frame into a compressed latent representation, allowing for realistic reconstruction. This method is efficient, making it ideal for video transmission applications where perceptual accuracy is more important than perfect pixel accuracy. These techniques enable efficient compression without significant quality loss, representing a paradigm shift in modern video compression techniques. After analyzing the results the highest compression ratio (94.75) was achieved by compressing using the GAN algorithm. The highest PSNR (36.08) was achieved by CNN+RNN(LSTM) compression.

1. Introduction

With the rapid development of artificial intelligence and computing science artificial neural networks have become one of the most prominent tools used in processing complex data especially in fields that require high accuracy and efficiency such as image and video analysis. These networks are based on models inspired by the human brain, and are notable for their ability to learn from data and extract significant patterns, making them suitable for a wide range of applications.

Neural networks play a growing role in numerous fields such as face recognition image quality enhancement object detection and data compression thanks to their adaptive capabilities and machine learning capability. They are widely used today in intelligent monitoring systems mobile applications and digital platforms where they contribute to improving efficiency reducing resource consumption and improving performance in multimedia analysis and processing [7].

Corresponding author E-mail address: dac2020@mtu.edu.iq
https://doi.org/10.71229/sqe4h444

The research problem explains Neural networks are a good way to compress videos but using them requires massive processing power and huge amounts of training data which increases processing time and resource consumption. The complexity of the embodiments may also require sophisticated hardware making it difficult to achieve a balance between compression ratio and video quality particularly in complex cases [7].

The research aims to the use of neural networks for video compression aims to reduce the size of files improving the user experience and speeding up online uploads and downloads. It also reduces the amount of data needed for transmission lowering internet costs. Moreover video compression makes it easier to store video files more efficiently by freeing up storage space on local devices and servers.

The importance of the research is in Neural networks are used in video compression to help reduce the size of files speeding up Internet uploads and downloads especially in situations that require high speed such as live streaming or on sites with slow connections. It also helps save storage space by enabling more videos to be stored on one volume. Especially in limited internet packages video compression reduces expenses by reducing data consumption. It also reduces clipping during live streaming improving the quality of the online viewing experience. video compression makes it easy to share files via email cloud and social media platforms that have file size limits. This results in faster and more efficient storage uploading and downloading.

2. Literature Review

Here are some literature reviews related to the topic:

A study by Ma, Siwei, Zhang, Zhang, et al.(2020) proposes a new method for video compression using neural networks specifically convolutional networks (CNNs) within the HEVC framework. The method is based on converting frames into compressed representations using convolutional neural networks then efficiently quantizing and reconstructing them. The model integrates in-frame prediction optimization motion compensation and filtering achieving high compression ratios while maintaining video quality[1].

A study by Mahmoud Darwish and Magdy Bayoumi in 2024 proposed a new model for video quality control and compression using neural networks combining convolutional networks (CNNs) and recurrent neural networks (RNNs). The algorithm is based on video feature extraction using convolutional neural networks (CNNs) followed by sequential per-frame bitrate predictions using RNNs enabling efficient video compression based on network conditions and user preferences. Results show that the model reduces reload rates by up to 87.5% improves quality of experience by 16.6% and reduces bandwidth usage by 37.1% making it suitable for online streaming applications such as DASH [2].

A study by Mustafa Shukur, Bharat Damodaran, et al, (2022). In this study a new model is proposed to compress facial videos using competitive generative networks (GANs) specifically StyleGAN2. Each frame of video is represented in the latent space of the model allowing effective frame compression without loss of visual quality. The algorithm is based on transforming the latent space into a new representation using a

normalization flow algorithm trained to improve compression efficiency and reduce perceptual distortion. Differential coding between frames is also applied in this space allowing compression inside and outside the frame. The results show that the proposed algorithm (SGANC) outperforms conventional compression algorithms such as VTM and AV1 in terms of optical quality at extremely low compression rates while maintaining highly realistic and noise-free images[3].

The current paper involves the use of two neural networks: Generative Adversarial

Networks (GANs) and Hybrid Neural Networks (CNN+RNN(LSTM)).

3. Methodology

3.1. Video compression

Video compression is a process that aims to reduce the size of video data by exploiting spatial and temporal redundancy within and across frames while preserving image quality as much as possible. Frames are divided into different categories (I-frames, P-frames, and B-frames) and are compressed either independently or in response to the expectations of other frames in traditional compression algorithms. Motion compensation used to depict the movement of objects between frames and residual compression resulting from the difference between the original frame and the modified frame are basic compression procedures. [4]

Video compression features include [1]:

• Minimize temporal and spatial redundancy which reduces the size of transmitted data and increases storage and transmission efficiency.

• Utilize intra- and inter-frame redundancy to achieve high compression rates.

• Reduce the bit rate required for video transmission to support capacity-constrained networks.

• Minimize loading and segmentation times to improve the streaming and viewing experience.

• Use neural networks to optimize performance enabling smarter motion adaptation and efficient compression.

Video compression reduces internet consumption speeds up transmission and reduces data size. It also improves the streaming experience by minimizing choppiness. Neural network technologies provide an adaptive distribution of data rates based on scene complexity optimizing video compression and image quality. [1]

3.2. Video File Format

Types of multimedia containers that store audio video and additional metadata are known as digital video file formats. Digital video data is stored on a computer system or digital device in a file type called video file format[5]. This file type contains container information metadata and encrypted streams. Video efficiency device compatibility video quality and size are all affected by the choice of format [6]. The most common formats are:

- AVI (Audio Video Interleave): A multimedia container developed by Microsoft in 1992 often used in digital cameras to save synchronized music and video files[6]. Data in this format is organized according to flexible device-specific structures (RIFF, hdrl, movi, idx1, trash). Forensic analysts can identify the source or detect changes using these structural differences such as the naming of codecs or the location of metadata[5].
- (MOV) Apple QuickTime Movie : Apple's multimedia container is used in mobile phones and digital cameras arranged in adaptive structures called atoms. It uses codecs like H.264 and MP4V distinct atomic layouts and metadata which help determine device location and identify video changes even in lossless scenarios[5] .
- MP4 (MPEG-4 Part 14) : It is a well-known multimedia container often used in editing and mobile applications. It supports encodings such as H.264 and MP4V and stores data in formats such as ftyp, moov, and mdat[6] .
- MPEG (Moving Picture Experts Group) : The WMV, MPEG, AVI, and MOV video formats are popular for long-term digital preservation due to their cross-platform compatibility efficient compression and support[6] .

3.3. Mathematical operation

A. Hybrid neural networks (CNN+RNN(LSTM))

The proposed hybrid technique achieves efficient video compression by combining the capabilities of recurrent neural networks (RNNs) and convolutional networks (CNNs) supported by long-term memory (LSTM) modules. Focusing on highlighting important visual information and removing unnecessary details RNNs create a denser representation of each video frame by extracting important spatial features. Redundant data within a single frame is minimized in this step [8].

Then a recurrent neural network (LSTM) receives the retrieved spatial information and examines the temporal correlations between subsequent frames. LSTM helps remove the temporal redundancy that often exists between closely spaced frames by capturing dynamic patterns across time such as motion and direction. The combination of spatial and temporal extraction results in highly efficient compression reducing the amount of data provided without significantly affecting the quality of the final video while preserving essential information about both visual structure and temporal changes [9]. Here are the mathematical formulas used:
CNN's representation of spatial extraction:

$$CNN(X_t) = F_t \quad \ldots(1)$$

Where:
X_t: It is the frame in time t .
F_t: It is the representation of spatial features.

LSTM for managing temporal dependence:

$$LSTM(F_t, h_{t-1}) = h_t \quad \ldots(2)$$

Where:
h_t :It is the representation of spatial and temporal features after combining them.
h_{t-1}: It is the previous hidden state (memory).

B. *Generative Adversarial Network (GAN)*

A deep learning system known as competitive generative networks (GANs) consists of two competing networks: a discriminating network that distinguishes between produced data and real data and a generating network that aims to generate real data. GANs are used in the context of video compression to reduce the amount of data transmitted while reconstructing video frames with excellent visual quality after compression. Unlike traditional methods that rely on classical processing techniques GANs improve efficiency and reduce bitrates while maintaining accurate and realistic features in compressed frames. [10].

How the algorithm works[11 ،10] :

- Reference frame compression using GAN: Reference frames are first encoded using a trained GAN to ensure that the compressed frames retain an optical quality close to the original frame.
- Inter-frame motion estimation: The optical flow between the reference and target frames is calculated using the Motion Estimation and Compensation module.
- Prediction frame generation: To reduce the amount of additional data needed for transmission a prediction frame is produced using a reference frame and motion compensation.
- Calculation of residual frame: To generate a residual frame containing only the unexpected information, the difference between the actual target frame and the expected frame is calculated.
- Using GAN to compress the remaining frame: The encoder compresses the remaining frame into properties. The frame is then reconstructed with high accuracy by running random characteristics and noise through a generator.
- Multi-scale frame review: To ensure that fine detail is maintained at all scales generated frames are evaluated with multiple resolutions.

- Training with sophisticated loss functions: To balance image quality and compression rate the technique uses a combination of distortion loss (MSE), generator loss, and feature matching loss.

Basic mathematical formulas:

$$L_{distortion}(E,G) = MES(X_t, \widehat{x_t}) \quad \ldots(3)$$

Where:

X_t : ground-truth target frame.

$\widehat{x_t}$: the decoded target frame.

Generator Loss:

$$L_{generator}(E,G) = \left\|\hat{F}^1_{1,5-} 1^1\right\|^2_F \quad \ldots(4)$$

Where:

$\hat{F}^1_{1,5}$: is the feature map extracted at layer 5 of the discriminator from the full-resolution decoded

1: is an all-one tensor of the same shape.

Feature Matching Loss:

$$F_{feature}(E,G) = \sum_{S=1,\frac{1}{2},\frac{1}{4}} \sum_{j=1}^{4} MSE(F^s_{t,j}, \hat{F}^s_{t,l}) \quad \ldots(5)$$

Where:

$F^s_{t,j}$: Features extracted from the real frame.

$\hat{F}^s_{t,l}$: Features extracted from the resulting frame for the same scale and layer.

Total loss of encoder and generator:

$$L_{encoder_{generator}}(E,G) = L_{generator}(E,G) + \lambda_x L_{distortion}(E,G) + \lambda_f F_{feature}(E,G) \quad \ldots(6)$$

Where:

λ_x: To weigh the distortion loss.

λ_f: To weigh the feature matching loss.

The algorithm relies on high-frequency recovery in the wave domain to improve the quality of the compressed video to a level that mimics the quality of the original video focusing on improving the visual experience rather than simply improving traditional quality metrics.

3.4 Performance Evaluation Metrics

The performance of the model was evaluated using common evaluation metrics which include compression ratio and signal-to-noise ratio (PSNR).

a. Compression Ratio

It is a key metric in video compression used to measure how small a video file is and determines the efficiency of the compression algorithm[4] . It is represented by the following formula:

$$Compression\ Percentage = \left(1 - \frac{Compressed\ Size}{Original\ Size}\right) \times 100\% \quad \ldots(7)$$

b. Peak signal-to-noise ratio (PSNR)

It is an objective measure commonly used to evaluate the quality of compressed images or videos. This rate measures the ratio between the maximum possible signal value (original image/video) and the power of the spoiling noise (difference after compression). It is usually expressed in decibels (dB). Higher PSNR values generally indicate better quality, that is, less distortion due to compression [12]. It is calculated using the following equation:

$$PSNR = 10.\log_{10}\left(\frac{R^2}{MSE}\right) \quad \ldots(8)$$

Where:

R: It is the maximum possible value for the pixel.

MSE: The mean squared error between the original image/video and the compressed image/video. MSE values are calculated using the following equation:

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \ldots(9)$$

Where:

I(i,j): is the original pixel value.

(i,): is the compressed pixel value.

m×n: is the image resolution.

3.5. Video Database

The video database (4) represents video files and the number of frames in each video varies based on the duration and frame rate of each clip. Videos vary greatly in duration, scene content and visual complexity ,Each video differs in terms of movement, whether fast or slow, lighting and edges. which contributes to the diversity of the database in different formats (MP4, MOV, and MPEG) divided into frames, with a total number of 3535 frames.
The input frames dimensions are standardized using a single resolution to satisfy the neural model's specifications and make processing and training easier.

Figure 1. One of Video data base

Table .1 Database

Video	Video duration	Video file format	Number of frames	Resolution
1	00:01:39	Mov	2995	432p (768×432)
2	00:00:04	Mp4	113	432p (768×432)
3	00:00:06	Mov	189	432p (768×432)
4	00:00:07	Mpeg	238	432p (768×432)

3.6. Video Compression Methods (VCM)

a. Video Compression using Hybrid Neural Network (CNN + RNN)

This method relies on video compression using a hybrid neural network (CNN + RNN). The model architecture can be explained through the following steps:

There are two primary steps to the model:

1. In the first step, visual features are extracted from each frame separately using a CNN. To minimize dimensionality, frames go through three to five consecutive convolutional layers (Conv2D), followed by pooling layers (Max Pooling). In order to convert the CNN outputs into compressed numerical representations, these layers usually end with a Flatten or Global Average Pooling operation.

2. The second step involves temporally ordering these extracted features and feeding them into a one- to two-layer LSTM network. This network comprehends the temporal linkages and sequence between frames, allowing the model to anticipate following frames or identify frame duplication. In certain configurations, the LSTM is followed by one or two fully linked (Dense) layers to carry out classification, prediction, or keyframe selection.

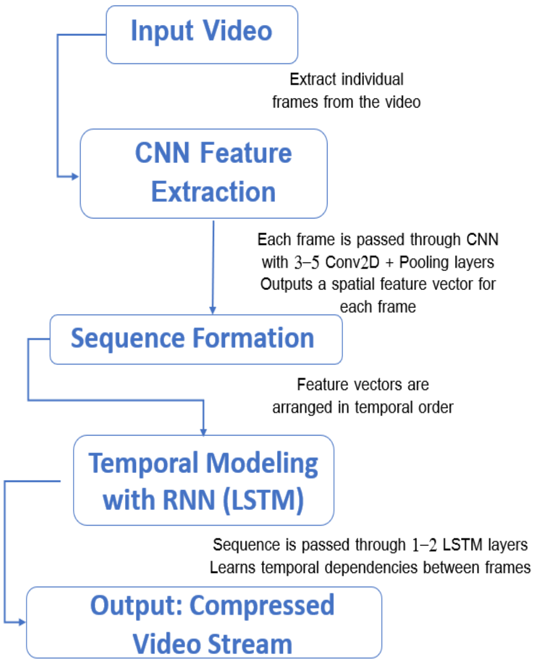

Figure 2. Schematic of CNN+RNN algorithm for video compression

b. Video Compression using Generative Adversarial Network (GAN)

This method relies on video compression using Generative Adversarial Network (GAN). The model architecture can be explained through the following steps:

1. Each video frame is first sent to a generator, which has four to six layers, including Dense and Conv2DTranspose. This layer takes a low-dimensional, compressed representation of the frame and turns it into a visual representation. To save space, this representation is kept rather of the entire frame.
2. After that, the frame is sent to a discriminator, which uses three to five Conv2D layers with LeakyReLU to identify if it is a real or regenerated frame.

The compression process gradually becomes better as the discriminator tries to identify phony frames and the generator learns how to produce high-quality frames during training. Only the generator is used to regenerate the video from the compressed representations after it has been retrieved.

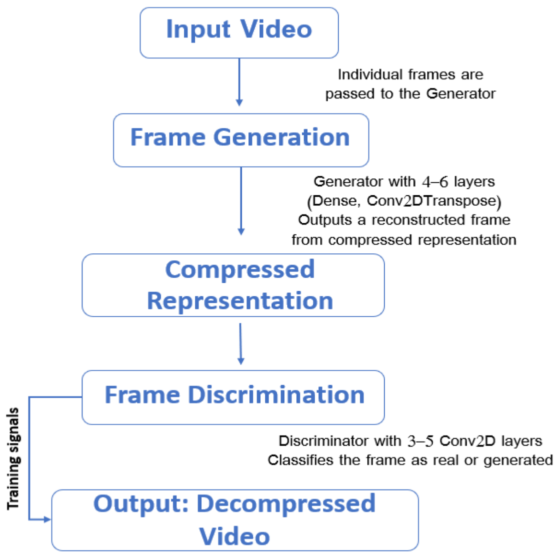

Figure 3. Schematic of GANs algorithm for video compression

4. Results and discussion

After applying the compressive methods on the video data base the following tables and fig. was shoed :

4.1 CNN+RNN(LSTM) algorithm results

Table . 2 CNN+RNN(LSTM) algorithm results

In	Video Name	Original Size (MB)	Compressed Size (MB)	Compression Ratio (%)	PSNR (dB)
1	1.mov	10.15	1.39	86.31	36.08
2	2.mp4	1.40	0.37	73.20	33.81
3	3.mov	3.44	0.49	85.89	33.33
4	5.mpeg	9.12	0.79	91.39	33.02

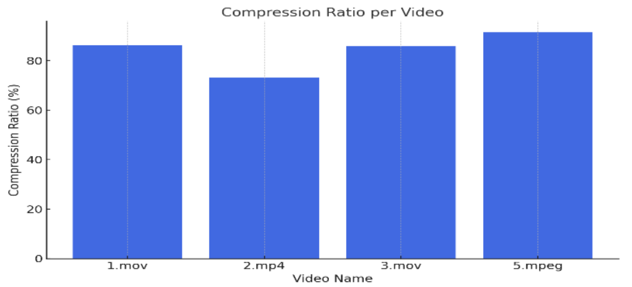

Figure 4. Compression ratio of CNN+RNN(LSTM) algorithm

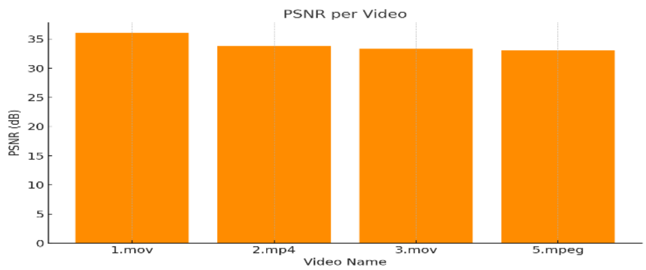

Figure 5. PSNR of CNN+RNN(LSTM) algorithm

Table .3 Overall results of CNN+RNN(LSTM) algorithm

Category	Original Size(MB)	Compressed Size(MB)	Compression Ratio%	PSNR
Average	6.02	0.76	84.19	34.06
Min	1.40	0.37	73.20	33.02
Max	10.15	1.39	91.39	36.08

From the above table, we notice that the average compression ratio for the Dataset was (84.19) the lowest size was (73.20) and the highest size was (91.39) and the average peak signal-to-noise ratio (PSNR) was (34.06) the lowest size was (33.02) and the highest size was (36.08) and we notice that compression depends on the type of video and the number of frames it produces from
its size index and we notice that it is affected by the general content of the video.

4.2 *GANs algorithm results*

Table .4 GANs algorithm results

In	Video Name	Original Size (MB)	Compressed Size (MB)	Compression Ratio (%)	PSNR (dB)
1	1.mov	10.15	1.18	88.35	24.07
2	2.mp4	1.40	0.40	71.49	24.38
3	3.mov	3.44	0.32	90.60	23.03
4	5.mpeg	9.12	0.48	94.75	20.00

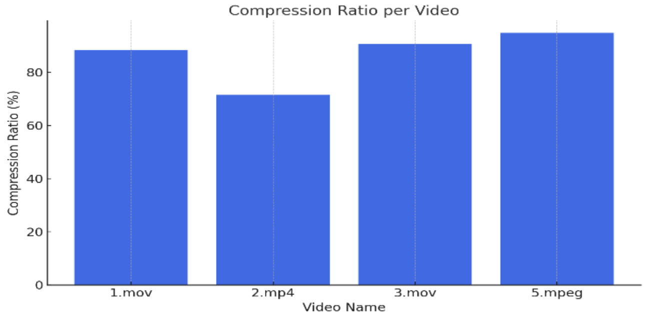

Figure 6. Compression ratio of GANs algorithm

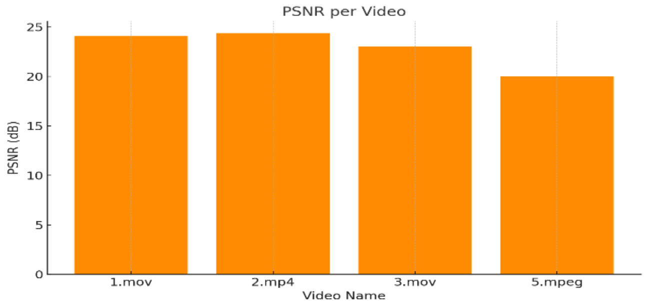

Figure 7. PSNR of GANs algorithm

Table .5 Overall results of GANs algorithm

Category	Original Size(MB)	Compressed Size(MB)	Compression Ratio%	PSNR
average	6.02	0.56	86.29	22.87
Min	1.40	0.32	71.49	20.00
Max	10.15	1.18	94.75	24.38

From the table above, we note that the average compression ratio for the database (Dataset) was (86.29) the lowest size was (71.49) the highest size was (94.75) the average peak signal-to-noise ratio (PSNR) was (22.87) and the lowest size was (20.00).The highest size was (24.38) and we note that the compression depends on the type of video As well as the number of frames it produces from its size indicator we notice that it is affected by the general content of the video.

5. Conclusions

After analyzing the results of the two algorithms according to compression ratio and PSNR several conclusions and recommendations stand out:

The best compression method with the highest compression ratio of 94.75 was achieved using the competitive generative network (GAN) algorithm.

The best compression method with the highest signal-to-noise ratio of 36.08 was achieved using a hybrid neural networks algorithm (CNN+RNN(LSTM)).

Suggested Recommendations: Use the GAN algorithm when the main goal is to reduce the video size as much as possible while taking into account some minor compromises in image quality especially in applications that require less storage capacity or faster network transmission.
Use the Hybrid CNN+RNN(LSTM) algorithm in scenarios where visual video quality is a higher priority, such as high-quality video streaming or virtual reality applications, even if the compression ratio is relatively lower.

References

[1] Ma, S., Zhang, X., Jia, C., Zhao, Z., Wang, S., & Wang, S. (2019). Image and video compression with neural networks: A review. *IEEE Transactions on Circuits and Systems for Video Technology*, *30*(6), 1683-1698.

[2] Darwich, M., & Bayoumi, M. (2024). Video quality adaptation using CNN and RNN models for cost-effective and scalable video streaming Services. *Cluster Computing*, *27*(5), 6355-6375.

[3] Shukor, M., Damodaran, B. B., Yao, X., & Hellier, P. (2022, October). Video coding using learned latent gan compression. In *Proceedings of the 30th ACM International Conference on Multimedia* (pp. 2239-2248).

[4] Rippel, O., Nair, S., Lew, C., Branson, S., Anderson, A. G., & Bourdev, L. (2019). Learned video compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 3454-3463).

[5] Krishnamurthy, C., & Angadi, M. (2014). Trends in Digital File Formats for Long-Term Preservation: An Overview. *Indian Journal of Library and Information Technology (IJLIT)*, *4*(2), 23-29.

[6] Laghari, A. A., He, H., Khan, A., & Karim, S. (2018). Impact of video file format on quality of experience (QoE) of multimedia content. *3D Research*, *9*, 1-11.

[7] Lam, Y. H., Zare, A., Cricri, F., Lainema, J., & Hannuksela, M. M. (2020, October). Efficient adaptation of neural network filter for video compression. In *Proceedings of the 28th ACM International Conference on Multimedia* (pp. 358-366).

[8] Ding, D., Ma, Z., Chen, D., Chen, Q., Liu, Z., & Zhu, F. (2021). Advances in video compression system using deep neural network: A review and case studies. *Proceedings of the IEEE*, *109*(9), 1494-1520.

[9] Bellantonio, M. (2016). *Hybrid CNN+ LSTM for face recognition in videos* (Master's thesis, Universitat Politècnica de Catalunya).

[10] Du, P., Liu, Y., Ling, N., Liu, L., Ren, Y., & Hsu, M. K. (2022). A generative adversarial network for video compression. *Proceedings of SPIE*, 129–136.

[11] Wang, J., Deng, X., Xu, M., Chen, C., & Song, Y. (2020, August). Multi-level wavelet-based generative adversarial network for perceptual quality enhancement of compressed video. In *European conference on computer vision* (pp. 405-421). Cham: Springer International Publishing.

[12] Deshpande, R. G., Ragha, L. L., & Sharma, S. K. (2018). Video quality assessment through PSNR estimation for different compression standards. *Indonesian Journal of Electrical Engineering and Computer Science*, *11*(3), 918-924.