



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 13, April 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Long Short-Term Memory Model based 3D Medical Image Compression

¹Dr.S.A.Yuvaraj, ²Amirtha Varshini M, ³Selvin B

Professor, Department of BME, GRT Institute of Engineering and Technology, Tiruttani, Tamil Nadu, India

UG Student, Department of BME, GRT Institute of Engineering and Technology, Tiruttani, Tamil Nadu, India

UG Student, Department of BME, GRT Institute of Engineering and Technology, Tiruttani, Tamil Nadu, India

ABSTRACT – Medical Image Analysis has become a major role in diagnostics operations in the healthcare industry and 3D image allows doctors to swiftly and accurately arrive at a conclusion on the effect of diseases on patients. These medical images are frequently shared with a concerned doctor or with experts in a particular medical field. Nowadays this transfer happens through wireless technology. So, it is vital to lower the data size of the medical image without any loss of information from it. This work proposes a technique to obtain a high compression ratio and fast encoding–decoding performance for the medical image by employing the optimal Long Short-Term Memory (LSTM) model. The proposed model is trained with 4-fold cross-validation on 12 high-resolution CT datasets while assessing the model's compression ratios and execution time. The experimental assessment was carried out on datasets gathered by different hospitals, representing diverse body segments, and unique scanning modalities (CT and MRI), and results are demonstrated improvements in comparison to the state-of-art system used for 3D-medical image compression.

KEYWORDS: 3D Medical Image Compression, Long Short-Term Memory (LSTM), Image Compression.

I. INTRODUCTION

Billions of medical images are created every year from various sources such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) (MRI). High-resolution images are crucial for diagnostics and clinical applications such as planning before the surgery, yet, significant storage needs (e.g., a 3D volumetric CT or MRI image might take a Gigabyte) can create a technological problem. These constraints impact not just limited storage space but also enhance the complexity of data flow and transmission. The effective compression methods can give answers to the aforesaid challenges with great reliability and precision without any loss of information from the image.

For predictive compression schemes, multiple local sample grids can be utilized. The input sequence form and size are critical in learning a mapping function from the input data of known causal neighbours to the desired output of the next unknown value for deep learning predictor-based models [1]. In general, there is a trade-off between the size of a model's sequence and its computing cost. Various block coverages around the target voxel can be applied to 3D data. This research investigates additional choices for the causal nearby sequence and analyses the size-speed compression trade-offs. In comparison to other recurrent unit alternatives, LSTMs are commonly used for tackling sequence and time series prediction issues because of their efficacy and training stability in handling vanishing gradient concerns. The LSTM algorithm was chosen because it effectively balances architectural compactness with the capture of long-term dependencies and correlations within spatial neighbourhood voxel sequences.

II. RELATED WORKS

To eliminate duplication, prediction-based compression is used, in which a decorrelation of the causal neighboring values is used to anticipate a target value. The approach is known as image-codec when compression is conducted over a 2D changing signal. A stack of 2D frames over time (video-codec) or a 3D changing signal might be termed 3D data (volume-codec). A residual error is commonly estimated after iteratively applying a compression model to build a map of predicted values to quantify the difference between forecasts and ground truth values. To eliminate coding redundancy, this prediction error will be reduced losslessly using an entropy coder, such as arithmetic coding, context adaptive binary arithmetic coding, Asymmetric Numeral Systems (ANS), or Huffman coding. Within the predictive-based lossless compression literature, many linear or non-linear combinations of causal neighborhood values,

the number of neighbors, and the form of their sampling patterns are used, encompassing both classical and learning-based prediction techniques.

A. Methods based on traditional prediction

The Joint Photographic Experts Group-Lossless (JPEGLS) uses the immediate three pixels neighborhood to forecast a target pixel using a mode-selection scheme using the LOCO-I algorithm, which is one of the classic predictor-based approaches in the image codec category [2]. Context-Based Adaptive Lossless Picture Codec (CALIC) is a more advanced approach for image compression that incorporates six pixels of causal information into a context-based Gradient Adjusted Predictor (GAP) [3]. The Minimum-Rate Predictor (MRP) is an adaptive predictive-based system that uses 2D-block categorization and has a larger causal neighborhood [4]. Several traditional image coders have expanded their capability to 3D space in order to compress higher dimensional volumetric data, such as films and 3D medical scans. Extended versions of the image coder CALIC are (3DCALIC) [5] and (M-CALIC)[6]. 3D-CALIC is an improved version that allows for context decorrelation in both interband and intraband modeling. The M-CALIC method, on the other hand, outperforms 3D-CALIC when it comes to decorrelating hyperspectral data with multiband lossless and near-lossless compression. For both 8 and 16-bit depth contents, 3D-MRP extends the MRP method to include 3D causal neighborhood pixels and delivers an improved error estimation and context estimator [7].

High-Efficiency Video Coding (HEVC)[8] is a well-known 3D codec for video compression that integrates many coding methods and offers lossy and lossless compression choices. The lossless mode is a predictive-based technique that reduces data redundancy inside and between frames by using both inter and intra-prediction. 3D medical imaging employing Range Extension [9] with 4:0:0 Chroma format for one channel component 16-bit data is one of the HEVC lossless mode applications. In conclusion, the majority of traditional lossless compression methods rely on hand-crafted or linear combinations with a few causal neighboring pixels coverage for their predictions. Furthermore, such algorithms are only meant to perform effectively on the data domains for which they were created, which are often natural pictures or video sequences.

These major drawbacks need the development of new methods for predicting non-linearity that is more flexible. Deep learning algorithms have a lot of promise and are a promising research topic since they can represent non-linear data distributions with a lot of efficacy and flexibility.

B. Learning-based prediction algorithms

Compared to conventional state-of-the-art compression approaches, the current deep learning algorithms are getting outstanding compression results for both lossy and lossless compression [10]. This trend towards the learning-based approaches is subsequent to its outstanding outcomes performance in many areas, amazing ability in modeling nonlinearity, and GPU utilization. Numerous learning-based approaches are proposed for lossy compression for applications including image-super resolution [11], generative compression [12] and end-to-end compression frameworks [13].

Sequence models are a special form of supervised learning algorithm that applies a prediction scheme. This neural network offers outstanding versatility for diverse sequential input and output of arbitrary lengths that a model may retain. Based on the length of the processed sequences, this model may be classed into numerous types, including one-to-one, many-to-one, one-to-many, and many-to many sequences prediction models. The areas for such models are numerous for sequential and higher-dimensional data, including sequence prediction, sequence creation, sequence categorization, and sequence-to-sequence prediction. A few examples of applications are weather predicting, product suggestion, stock market prediction, sentiment analysis, text translation, picture caption, and text generation[14]. Some state-of-the-art sequence prediction models have an internal state or memory unit, which helps to learn the long-term temporal contextual information. Generally, when solving a sequence prediction issue, a model learns a mapping function $f(xt)$, which transfers an input series xt to an output sequence yt . Examples of the state-of-the-art deep learning sequence prediction models are Recurrent Neural Network (RNN) [15], Long Short-Term Memory (LSTM)[16].

A more recent lossless compression method that leverages a CNN as a predictor for video coding was produced in [17]. This block-wise compression approach replaced all intra-prediction modes of the HEVC with deep learning CNNs and obtained an average bit reduction of 5 percent compared to the standard HEVC. Another deep learning lossless compressors that apply predictive strategy but for sequential data include Deep Zip [18], LSTM-Compress [19], and Dzip [20] in these compression frameworks, a mix of neural network-based compressor and

arithmetic coding is applied. Both Deep Zip and LSTM-Compress involve RNN-based models (e.g. GRU or LSTM models) as probability estimators together with an arithmetic coding unit and especially employed for losslessly compressing text and Genomic datasets. DZip is a more general-purpose NN-based model for reducing multiple dataset types utilizing a hybrid training technique.

III. METHODOLOGY

A. Problem description

We select numerous causal adjacent sequences X_n for training from a data distribution specified over $V \subset \mathbb{R}^N$, where $X_n \in V$. As shown in Fig. 1, each sequence X_i has a series of observations of surrounding adjacent voxels' intensities $X_i = x_0, x_1, x_2 \dots x_{l-1}$ with a fixed sequence length l . The LSTM model should train a differentiable mapping function $\hat{y}_i = f(X_i)$, which transfers a series of intensity values X_i to a prediction of the next single target voxel value \hat{y}_i . The LSTM predictor model learns to minimize the difference between the prediction value \hat{y}_i and the ground truth value y_i by backpropagating during training. The residual or prediction error $E = y_i - \hat{y}_i$ is computed while evaluating the model. The volumetric prediction error is then losslessly reduced to a lower bit rate using an arithmetic cosine transform. An arithmetic decoding is used to decompress the volumetric error on the receiver side in order to retrieve the original volume. The LSTM model will then auto-regressively create the prediction values \hat{y}_i , which are subsequently added to the residuals E .

B. Causal Neighboring Sequence

The input sequence is critical in learning the mapping function to the goal output in a predictive-based model. There is, of course, a trade-off between the quantity of information a sequence provides the LSTM model and the computing cost. The longer the sequence, the slower the model develops, and the less stable the training becomes, the more likely it is to encounter gradient difficulties. This research looks at and provides several approaches for encircling neighboring sequences generated from 3D high-resolution medical volumes. As seen in Fig. 2, two alternative block forms were introduced: the cube and the pyramid. Several sizes (distances to the target voxel) were applied to each form. We tested voxel sequences with 1D, 2D, and 3D coverage

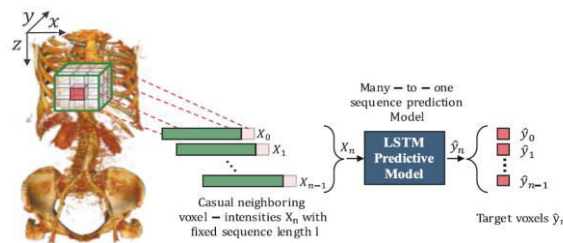


Fig 1. Illustration of the supervised learning LSTM model with an explicit overview of the method for extracting the causal neighbouring sequence from 3D medical images (16 bit-depth).

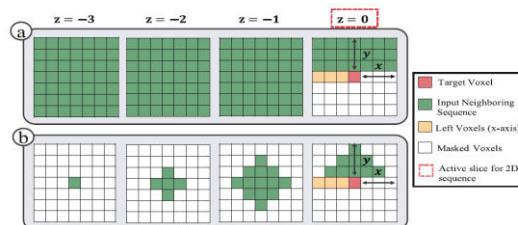


Fig2. A demonstration of the proposed causal neighbouring sequences that can have different shapes, dimensions, block sizes, and sequence lengths. The two main shapes are (a) Cube and (b) Pyramid.

The red voxel refers to the target voxel to be predicted while the green voxels form the model's input sequence and the white voxels are masked (excluded). $z = 0$ is the current slice which is also the only active slice when extracting a 2D sequence.

C. Model

The lossless compression task is formulated as a supervised machine learning assignment. The suggested approach is also known as a many-to-one prediction model. Since the LSTM model learns a mapping from the input sequence to the intended output via the backpropagation process. The LSTM model was chosen since it is expressly developed for

sequence prediction issues and can handle gradient difficulties (e.g. exploding or disappearing gradients) better than the RNN because to its gating mechanism. These characteristics enable LSTM to develop long-term reliance with more consistent training. The input list is crucial in learning the mapping to the objective output for a sequence prediction model. A sequence is defined as an ordered series of observations that travel through the hidden cells of an LSTM in a specific order.

D. Hyperparameters training

The suggested model is made up of a single 128-unit LSTM layer, followed by a linear output layer, which serves as the basic architecture and has the same weights for both the encoder and decoder. The model's weights need just 276 KB of storage, however, the entire model (weights and training hyperparameters) takes about 810 KB (KB). All of the suggested models are optimized using the Adam optimizer with parameters $\beta_1 = 0.9$, $\beta_2 = 0.98$, $1e-4$ learning rate, and 128 batch size.

$$L_{\text{joint}} = \text{MAE} + (\lambda |\text{PCC}|) \tag{1}$$

$$\text{MAE} = \sum_{i=1}^n |y_i - \hat{y}_i| / n \tag{2}$$

and the Pearson Correlation Coefficient (PCC).

$$\text{PCC} = \text{cov}(y_i, \hat{y}_i) / \sigma_{y_i} \sigma_{\hat{y}_i} \tag{3}$$

where y_i is the ground truth voxel value, \hat{y}_i is the model's prediction, n is the total number of data samples, cov is the covariance, σ_{y_i} , and $\sigma_{\hat{y}_i}$ are the standard deviations of y_i and \hat{y}_i . Because PCC assesses the statistical connection between two continuous variables, we combine it with MAE to improve both accuracy and stability when solving regression problems.

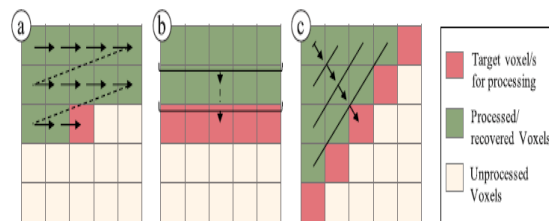


Fig3. The order at which the decoder will decode voxels based on the defined causal neighbourhood specifications. (a) The order of decoding voxels (sequentially) restricted by neighbourhood dependencies applied in 3.4.1. (b) The order of decoding voxels (in parallel) by decoding an entire batch applied in 3.4.2. (c) a possible parallel approach to decoding voxels, which does not allow full GPU occupancy.

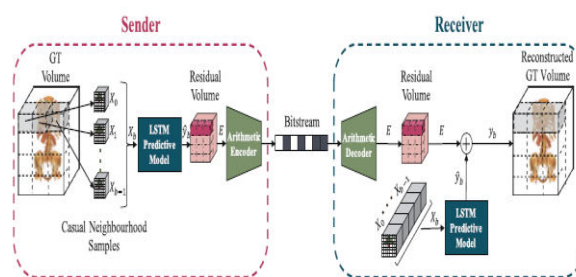


Fig4. Overview of the optimised version of our lossless compression framework using LSTM.

This version leverages parallelism by employing a novel input sequence, which adds more flexibility and allows the decoder to process several batches of sequences in parallel

IV. RESULTS

A. Dataset

The first dataset utilized in this investigation is Dataset1, which consists of 12 high-resolution CT private volumes of the human torso provided by a local hospital. This collection includes 16-bit greyscale Digital Imaging and Communications in Medicine (DICOM) photos for each volume. All scans have a resolution of 512 512 pixels, a pixel spacing of [.488, .488], and a slice thickness of.625 mm. The number of photos per volume varies $z \in [728, 1008]$. In all volumes, the least intensity value is 1024, and the maximum is 3071. Figures 4(a) and 4(b) depict orthogonal slice views and 3D volume renderings of two sample volumes from Dataset1 (b).A public dataset known as Dataset2 was

used solely for the purpose of performing experimental assessment. The DICOM files in Dataset2 make up a total of 12 MRI volumes of patients' head and neck scans. Each picture has a resolution of 512 X 512 pixels, [.5, .5] pixel spacing, and a slice thickness of 2 mm. There are 120 frames each volume, with a minimum intensity value of 0 and a maximum intensity value of 689. In Fig. 4, orthogonal slice views of one sample volume from Dataset2 are depicted (c).

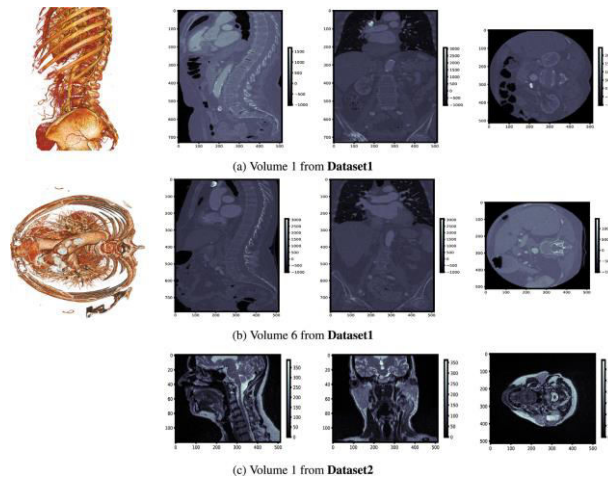


Fig. 5(a) and (b) illustrates 3D volume visualisation of patient’s entire Torso with three orthogonal slice views (axial, sagittal and coronal) of two sample volumes from Dataset1.

While Fig. 5(c) presents three orthogonal slice views (axial, sagittal and coronal) of a single sample MRI volume from Dataset2 illustrating the patient’s head and neck.

B. Performance metrics

Over Dataset1, we used 4-fold cross-validation to assess the accuracy and performance of our LSTM prediction models for each of the different neighboring sequence types. There are nine volumes in the training set and three in the testing set for each fold. In general, a validation procedure like this is used to assess the efficacy and generalization of a trained model on unknown data. Furthermore, further experimental tests were carried out on an out-of-domain public dataset Dataset2 to assess the generalizability and resilience of the trained models across a variety of previously unexplored and unique medical modalities (e.g. MRI). The bits-per-pixel (bpp) (4) has been chosen as the assessment metric for all of our LSTM models. For both the encoding and decoding processes, compression time (in seconds) is monitored per model.

$$bpp = \text{Compressed Image Size (Bits)} / \text{Number of Voxels} \quad (4)$$

Figure 6 shows a summary of the average storage impact of the technique bpp over all 12 volumes for the two tests, whereas Figure 7 shows the change in (bpp) over-training epochs while compressing the same volume. When training models using varied neighborhood sequences, such as the pyramid input vectors, the compression ratio increases in this experiment (including left voxels).

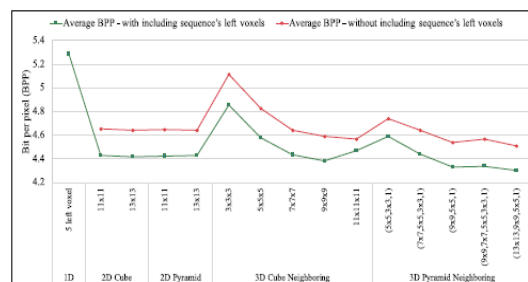


Fig 6. A summary overview of the average (bpp) over Dataset1’s volumes using all models trained with left voxels (in Green) and without (in Red) applied respectively. Overall, models pre-trained on reduced inputs result in a negligible performance drop of ≈ 0.2 bpp, but with a significant positive impact on the compression times

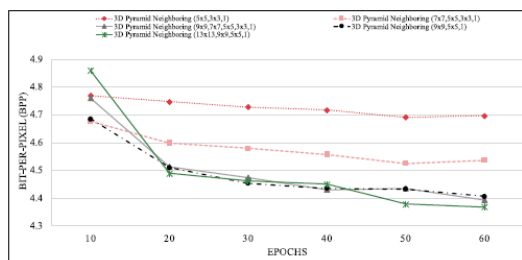


Fig7. Evaluation of trained models' compression qualities after different epochs. Models with pyramid input vectors are evaluated after 10, 20, 30, 40, 50, and 60 epochs.

V. CONCLUSION

A supervised LSTM prediction-based model for losslessly compressing 3D medical pictures (12-bit data saved as 16-bit depth volumes) is presented. As major assessment measures, we used comparison performance (bpp and time in seconds). Furthermore, various causal neighborhood sequences were empirically studied and analyzed to highlight trade-offs. Furthermore, immediate left voxels), which enables for easy decoder parallelism while maintaining a good compression ratio. We conclude from the experimental results that the pyramid-shaped input sequences achieve state-of-the-art compression results when compared to the other options, with a compression gain of 17 percent and 12 percent over Dataset1 and Dataset2, respectively when compared to the classical lossless compression alternatives.

The balance between compactness and representativity, which reflects the local correlation surrounding a target voxel, is what drives its efficacy. Furthermore, in the best scenario, the suggested shortened sequences achieve almost a 3 lossless compression ratio of the original volumes, and up to 500 decoding speed-up may be realized with minimal impact on compression performance. Furthermore, the proposed trained models outperform other state-of-the-art lossless codecs in compressing all 3D medical volumes and for different modality types, including CT and MRI, compared to other state-of-the-art lossless codecs. We intend to expand our work in the future to incorporate various types of deep learning NN models for a predictive-based compression domain. Another intriguing research topic is to look into the generalization of deep learning models to additional high-dimensional domain applications like video.

REFERENCES

- [1]. Nagoor, Omniah, Joss Whittle, Jingjing Deng, Benjamin Mora, and Mark Jones. "Sampling Strategies for Learning-based 3D Medical Image Compression." *Machine Learning With Applications* 8 (2022): 100273.
- [2]. Weinberger, M. J., Seroussi, G., & Sapiro, G. (2000). The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS. *IEEE Transactions on Image Processing*, 9(8), 1309–1324.
- [3]. Wu, X., & Memon, N. D. (1996). CALIC—a context-based adaptive lossless image codec. In 1996 IEEE international conference on acoustics, speech, and signal processing conference proceedings (pp. 1890–1893). IEEE Computer Society.
- [4]. Matsuda, I., Mori, H., & Itoh, S. (2000). Lossless coding of still images using minimum rate predictors. In *Proceedings of the 2000 international conference on image processing* (pp. 132–135). IEEE.
- [5]. Wu, X., & Memon, N. D. (2000). Context-based lossless interband compression extending CALIC. *IEEE Transactions on Image Processing*, 9(6), 994–1001.
- [6]. Magli, E., Olmo, G., & Quacchio, E. (2004). Optimized onboard lossless and near-lossless compression of hyperspectral data using CALIC. *IEEE -Geoscience and Remote Sensing Letters*, 1(1), 21–25.
- [7]. Lucas, L. F. R., Rodrigues, N. M. M., da Silva Cruz, L. A., & de Faria, S. M. M. (2017). Lossless compression of medical images using 3-D predictors. *IEEE Transactions on Medical Imaging*, 36(11), 2250–2260.
- [8]. Sullivan, G. J., Ohm, J., Han, W., & Wiegand, T. (2012). Overview of the high-efficiency video coding (HEVC) standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(12), 1649–1668.
- [9]. Bossen, F., Flynn, D., Sharman, K., & Sühring, K. (2019). HEVC format range extension (RExt). Fraunhofer Heinrich Hertz Institute. URL <https://hevc.hhi.fraunhofer.de/rext>.
- [10]. Hussain, A. J., Al-Fayadh, A., & Radi, N. (2018). Image compression techniques: A survey in lossless and lossy algorithms. *Neurocomputing*, 300, 44–69.
- [11]. Lai, W., Huang, J., Ahuja, N., & Yang, M. (2017). Deep Laplacian pyramid networks for fast and accurate super-resolution. In *2017 IEEE conference on computer vision and pattern recognition* (pp. 5835–5843). IEEE Computer Society.



Organized by

Dhaanish Ahmed Institute of Technology, KG Chavadi, Coimbatore, India

- [12]. Santurkar, S., Budden, D. M., & Shavit, N. (2018). Generative compression. In 2018 Picture coding symposium (pp. 258–262). IEEE, <http://dx.doi.org/10.1109/PCS.2018.8456298>.
- [13]. Hu, Y., Yang, W., Ma, Z., & Liu, J. (2020). Learning end-to-end lossy image compression: A benchmark. CoRR, abs/2002.03711. URL <https://arxiv.org/abs/2002.03711>. arXiv:2002.03711.
- [14]. Brownlee, J. (2017). Long short-term memory networks with python: develop sequence prediction models with deep learning. URL <https://books.google.com.sa/books?id=ONpdsWEACAAJ>.
- [15]. Mikolov, T., Karafiát, M., Burget, L., Cernocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. In T. Kobayashi, K. Hirose, & S. Nakamura(Eds.), INTERSPEECH 2010, 11th Annual conference of the international speech communication association (pp. 1045–1048). ISCA.
- [16]. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- [17]. Schiopus, I., Huang, H., & Munteanu, A. (2020). CNN-based intra-prediction for lossless HEVC. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(7), 1816–1828.
- [18]. Goyal, M., Tatwawadi, K., Chandak, S., & Ochoa, I. (2019). DeepZip: Lossless data compression using recurrent neural networks. In A. Bilgin, M. W. Marcellin, J. Serra-Sagristà, & J. A. Storer (Eds.), *Data compression conference* (p. 575). IEEE.
- [19]. Knoll, B. (2020). lstm-compress: data compression using lstm. URL <https://github.com/byronknoll/lstm-compress>.
- [20]. Goyal, M., Tatwawadi, K., Chandak, S., & Ochoa, I. (2020). DZip: Improved generalpurpose lossless compression based on novel neural network modeling. In A. Bilgin, M. W. Marcellin, J. Serra-Sagristà, & J. A. Storer (Eds.), *Data compression conference* (p. 372). IEEE.



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details