Deep CNN Based Video Compression With Lung Ultrasound Sample

Helen. K. Joy[1*] and Manjunath R Kounte[2]

[1] School of Electronics and Communication Engineering, REVA University, Bengaluru, 560064, India
[2] Department of Electronics and Computer Engineering, School of ECE, REVA University, Bengaluru, 560064, India
[*] Corresponding author. E-mail: helenjoy88@gmail.com

Received: Nov. 02, 2021; Accepted: May 05, 2022

Video compression and transmission is an ever-growing area of research with continuous development in both software and hardware domain, especially when it comes to medical field. Lung ultra sound (LUS) is identified as one of the best, inexpensive and harmless option to identify various lung disorders including COVID-19. The paper proposes a model to compress and transfer the LUS sample with high quality and less encoding time than the existing models. Deep convolutional neural network is exploited to work on this, as it focusses on content, more than pixels. Here two deep convolutional neural networks, ie, P(prediction)-net and B(bi-directional)-net model are proposed that takes the input as Prediction, Bidirectional frame of existing Group of Pictures and learn. The network is trained with data set of lung ultrasound sample. The trained network is validated to predict the P, B frame from the GOP. The result is evaluated with 23 raw videos and compared with existing video compression techniques. This also shows that deep learning methods might be a worthwhile endeavor not only for COVID-19, but also in general for lung pathologies. The graph shows that the model outperforms the replacement of block-based prediction algorithm in existing video compression with P-net, B-net for lower bit rates.

Keywords: CNN, Motion estimation, COVID-19, P-frame, B-frame

http://dx.doi.org/10.6180/jase.202303_26(3).0002

1. Introduction to inter prediction

Inter prediction in video compression [1] provides better reduction in data by comparing reference from frames unidirectionally and bidirectionally, standard video codecs use the goodness of spatial and temporal redundancies in the frames to perform comparison. HEVC uses three different video frames to perform inter-prediction [2], I-frame or infra frame, P-frame or predicted frame, B- frame or bidirectional frame. Intra frame / I-frame is originally from the frame whereas P and B frame are predicted unidirectionally and bidirectionally, holds less bitrate to transfer but can deliver more information. The pixel movement in the video frame is defined using motion vector from one frame to another. The Fig. 1 shows the essence of inter prediction with motion estimation. In a particular frame with many coding tree units (CTU), the similarity index between the neighboring frames will be high mostly coz of the frame rate and continuity factor. The changes in the frames can be evaluated and send while encoding than the frame as it is or its similarities. This helps in better compression. By comparing the frames ahead and before and with its motion vector information the current CTU came be predicted. Within the search range (notated in doted area) the best match and its shift can be identified. This information helps in the best motion estimation in between the frames in HEVC.

Inter prediction is important in compression as the quality and size of raw video is more as the video acquisition device are getting advanced [3] and the resolution is increasing to 4K, 8K range. The number of pixels representing a frame holds more information, so compression in a need. The information correlation between the frames are

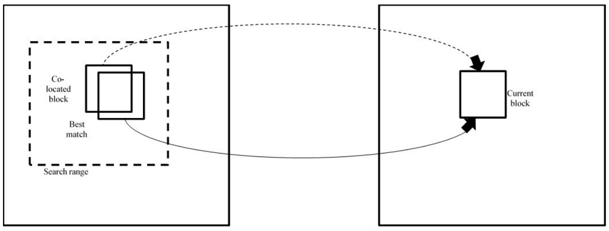

Fig. 1. Motion estimation to identify the best match in inter prediction in HEVC.

really high and there lies the hope of inter prediction, if the redundant information is identified and sent, the data rate will be less and can maintain better quality. The need of inter prediction mainly lies on the frame similarities or high frame correlation. Utilizing this, HEVC uses motion vectors and I, P and B frames to identify the redundant information and send to decoder with less information, I frame is extracted from raw frame and is transmitted directly. The P- frame or prediction frame is derived from the past frame and are send using optical flow vector. The B-frame or the bidirectional frame is derived by interpolating the previous frame and future frame and are sent using dense optical flow vector that are highly compressible. The motion vector delivers the information of pixel or block shifting [4] from one frame to another frame.

The dense optical flow [5] creates too many motion vectors that makes it complex. The shift to block-based motion compensation evolved here. The similar groups of pixels called macroblocks are transferred with motion information [6] thus resulting in high computational complexity. I-frame will be transmitted initially followed by the details of motion vector. The block-based motion compensation results in block artifact and it prefer only sequential decoding. Considering these cons, the deep learning came up with its advantage of content analysis to provide better solution for this.

The CNN networks are used to predict the P and B frame from the GOP [7]. The CNN encoder does the compression and binarization followed by decoding here the conditioning network provide information and decoder predicts the P and B frame. This approach outperforms the existing techniques for lower bit rate. Inter prediction in HEVC.

Inter prediction/motion compensated prediction is used to remove the redundant information from sending it to avoid usage of bits. This uses the difference between the best match object and its motion vector. It's performed on the Y or luminance component in YUV/YCbCr. Motion compensation [8] and motion estimation are part of it, where motion compensation is used in decoder to reconstruct the video frame and motion estimation used to identify the motion vector. The motion estimation (ME) [9] identifier the best match from future/ previous frame and the pixel movements will be mostly adjacent especially in high frame rate vector. The decoder receives the residual picture and the motion vector and it reconstruct the video frame [10] by motion compensation.

Both unidirectional and bidirectional prediction is done from the reference POC (picture order count) as it holds the predicted picture block of various frame. It uses the I, P and B frames to transfer the details in a particular order with I frame as reference frame and P&B followed by it. It follows some format with I and P with B in between usually.

2. Motivation and background for the proposed model

2.1. Background for the proposed model

Inter prediction is done with block partitioning method and comparing the similarity indexes between the blocks and the change in temporal domain by using motion vector in existing video compression scenario. HEVC uses inter frame, prediction frame and bidirectional frame to send the information after prediction. The P-frames are predicted with the help of I frame and B-frames with both I and p frame. This traditional block-based prediction was replaced by new researches focusing on fractional pixel method, virtual reference frame method, video prediction networks, convolutional neural network, recurrent neural network, deep learning etc.

To alleviate the complexity of block-based prediction, a CNN-based approach for fractional-pixel motion compensation to aggravate the efficiency of video coding was proposed [10]. Here, instead of interpolation the novel approach introduces the fractional-pixel MC as a regression problem between frames. A CNN was proposed to solve the problem, regression. The proposed method introduces a CNN models to generate a relationship between integer-pixel values and fractional references, each model in that accounts for a proper and fractional motion vector for bi-directional motion compensation the CNN has been trained properly.

Apart from the these, virtual reference frame generation (VRF), and enhanced motion-compensated [11] video coding using deep virtual reference frame (DVRF) [12] generation method that makes use of the reconstructed bi-directional frames to create a better efficient prediction of the to-be-coded frame was next stage in prediction. Developing a CTU level coding mode DVRF, that avoids the extra ME procedure on VRF will help to achieve a better trade-off between complexity and coding efficiency [13, 14].

Research have extended its idea to incorporate deep learning and neural network to extract the efficient version of inter prediction. Many researches happened in this with motion compensation model using CNN (CNNMCR) is CNN based motion compensation refinement that uses spatial correlation. Fractional pixel motion compensation with CNN was [15] introduced to create a CNN network to make a relationship between pixel values and fractional reference pixel. Virtual reference frame concept [16, 17] was introduced later that reduces the effect of compression artifact that focus on bidirectional frame reconstruction. The area for advancement is still open in this area. To use the deep learning to obtain the frames is the method followed here in the proposed method for P-frame and B-frame.

2.2. Motivation for proposed model

Lung ultrasound imaging is better compared to X-ray and CT scan because it is portable, harmless from radiations and inexpensive and can identify many lung diseases and it can be adapted for pregnant and people of old age group. As the recent studies shows that the LUS can be used to identify covid-19 [18] than chest radiograph (CXR) as shown in Table 1, X -ray [19], and its transfer and analysis are also an area to be focused. Out of the various techniques HEVC is preferred to be good for live transfer of videos, improving HEVC's prediction methods to transfer LUS video frames is the concern here as it will be harmless for human body. This is the factor drives in improvisation of HEVC prediction with CNN based P-frames and B-frames with the dataset with lung ultrasound videos. The transfer of lung ultrasound in an efficient mode is proposed in this paper. This is achieved with the help of CNN based prediction and bidirectional frame prediction method. Incorporating this idea and analyzing it with LUS dataset opens a vast hope of research in this area.

The proposed method focuses on predicting P and B frame using CNN network. The method proposed is briefly explained in the Fig. 2. The idea is to frame a CNN based encoder, that will encode and binarize the motion occurring in the group of pictures (GOP). Thresholding is done in CNN based encoder within the limit of {-1, 1} by adding quantization noise in uniform quantity. Decoder performs feature extraction transformation I-Frame. The conditioning network performs feature extraction of I- frame from existing image coded. The required feature is added to predict P&B frame in decoder correctly. The network learns and extracts the features from the given training inputs and predicts the frames with maximum efficiency.

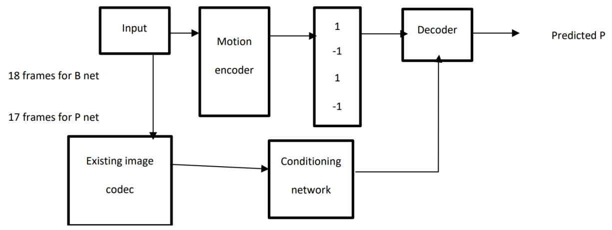

Fig. 2. Model of video prediction network. The I(intraframe), P (prediction frame) and B (bidirectional frame) are the input given to the network and the network predicts the P and B frame with learning based Deep CNN P and B net.

Table 1. Comparison of various features of lung ultrasound and chest radiograph [18].

Sl no		value	95%CI
	1.Sensitivity (%)		
1.1.	Lung ultrasound	88.9	71.1 – 97.0
1.2.	Chest radiograph	56.3	33.2 – 76.9
	2.Specificity (%)		
2.1.	Lung ultrasound	51.9	34.0 – 69.3
2.2.	Chest radiograph	75.0	50.0 – 90.3
	3.Positive predictive value (%)		
3.1.	Lung ultrasound	77.4	59.9 – 88.9
3.2.	Chest radiograph	77.8	54.3 – 91.5
	4.Negative predictive value (%)		
3.1.	Lung ultrasound	75.0	46.2 – 91.7
3.2.	Chest radiograph	48.0	30.0 – 66.5

3. Dataset for the proposed model

The dataset consists of lung ultrasound images and videos. The dataset combines data from collaborating hospitals as well as publicly available resources from the web. The LUS dataset consists of videos and images of various ultrasound of healthy, and affected lung ultrasound data. The details of LUS data used are provided below. The data set is growing by the contribution from medical fields.

The Convex probe has

- 162 videos (46x COVID, 49x bacterial pneumonia, 64x healthy, 3x viral pneumonia).
- 20 videos from the Butterfly dataset (18 COVID, 2 healthy).
- 53 images (18x COVID, 20x bacterial pneumonia, 15x healthy).

Linear probe has

- 20 videos (6x COVID, 2x bacterial pneumonia, 9x healthy, 3x viral pneumonia).

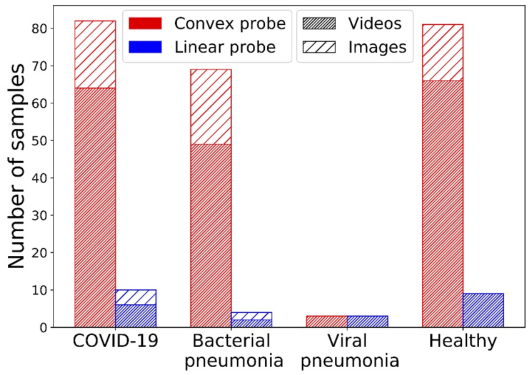

Fig. 3. Graphical representation of LUS dataset consisting of various lung ultrasound data of healthy and affected lungs.

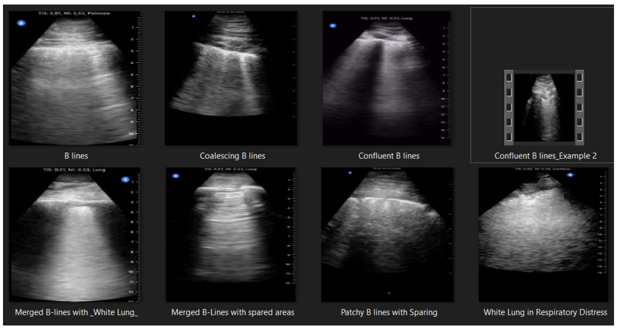

Fig. 4. Samples of LUS dataset is provided for reference.

- 6 images (4x COVID, 2x bacterial pneumonia).
- 45 videos of possible COVID-19 patients collected in Piacenza at the peak of the crisis in Italy; there were not enough PCR tests available, so the label is not clear.

The details are given in the graph shown in Fig. 3. The test to train data is 185:75. It shows the videos and images taken by convex and linear probe of ultrasound equipment; some sample data are provided in Fig. 4.

4. Design and analysis of prediction network (p-network)

P-net or prediction net is designed to predict the P-frame. The prediction frame is derived from the previous frame in one direction. Considering the prediction of P-frame using CNN in Fig. 5, the input used are previous frame and prediction frame of exiting data. The Eq. (1) shows how to derive the prediction frame from I_0. The traditional calculation was tedious in this case so the idea of CNN based prediction is a better idea as its efficiency proved was better in predicting the frame. The series of frame can be predicted by this network and its trained. The trained model was evaluated by LUS data set. The idea behind prediction is to use the CNN network with I&P as input with the details provided by the conditioning network. P-frame can be derived by the following equation

$$\begin{aligned}
\hat{P}_1 &= D[E(I_0, P_1)cond(I_0)] \\
\hat{P}_2 &= D[E(I_0, P_2)cond(I_0)] \\
\hat{P}_3 &= D[E(I_0, P_3)cond(I_0)] \\
&\;\;. \\
&\;\;. \\
\hat{P}_n &= D[E(I_0, P_n)cond(I_0)] \\
\hat{P}_x &= D\,[E(I_0, P_x)cond(I_0)]\; \forall x = 1, 2, ..., n
\end{aligned} \tag{1}$$

Here in Eq. (1), $\hat{P}_1, \hat{P}_2, ..., \hat{P}_x$ is predicted p frame or output, D is decoder, E represents encoder, I_0 is the input in the 0^{th} frame that is the previous intraframe or I frame and current $P_1, P_2, ..., P_x$ represents the current perdition frame.

The $\hat{P}_1$ (predicted p frame) is derived by the D(decoder) output of the E(encoder) output of I_0 (the input in the 0^{th} frame that is the previous intraframe or I frame) and current P_1 (current perdition frame) multiplied with conditioned I_0.

The GOP is the input of the P-net and its given parallelly to 3D and 2D convolution with kernels of different size to extract the global and local features correctly. The normalized 3D convolution output and 2D convolution is send to a series of convolution and pixel shuffle with shuffle factor of 2 to crunch the data. Finally, it provides a prediction frame from the given group of pictures of I and P frame.

The LUS dataset is used to evaluate the P-frame net. The variation in the kernel size helps the perdition precise the convolution in three dimensional and two dimensional helps the data extraction and prediction of P frames in more efficient and less complex way that makes the model better. The prediction of LUS frames with the p frame network makes the network to respond fast and helps from traditional calculation complexities. Even through the network design look tedious the prediction process is fast. The training process is long but the network performs fast once it is trained with optimum loss. Here the trained network shows and better output when evaluated with lung ultrasound videos. This procedure helps in its transfer and live telecasting for more analysis.

5. Design and analysis of bidirectional network (b-network)

B-network is a CNN based prediction network that uses the I and It+1 frame to predict the B-frame. The input given

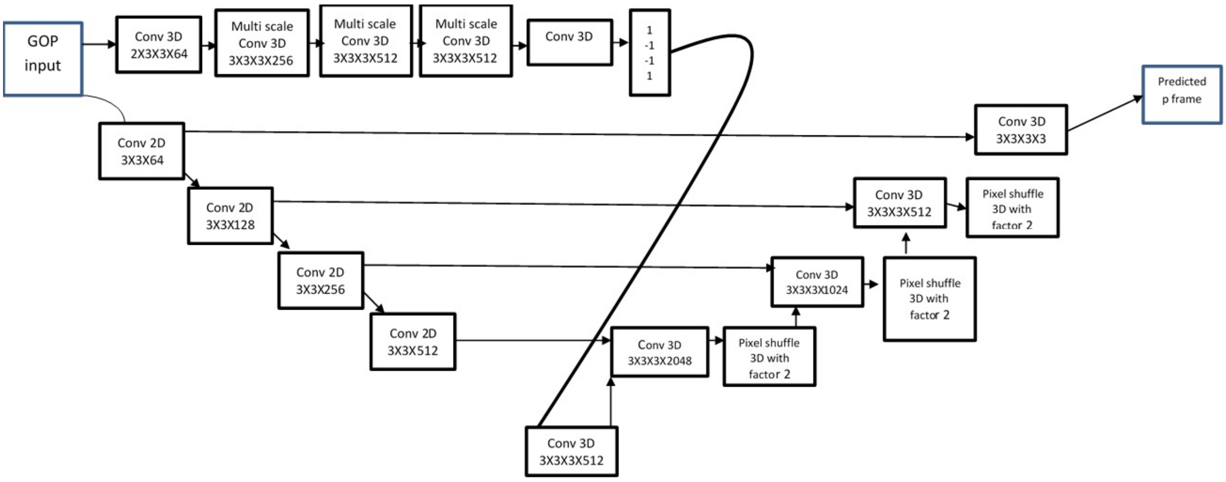

Fig. 5. The model of Deep CNN P-net.

to the binary motion encoder is the original GOP sequence, from hat the motion encoder uses the I-frame for conditioning network and other frame or B-frame to motion encoder as shown in Fig. 6. It uses supervised binary motion encoder with CNN to binarize the B-frame and provide the information along with conditioning network out to the decoder side of the thresholding operation.

The decoder uses the information from I-frame and the binarized output and shifts the pixels based on the information to predict the B-frame. The total operation of B-frame prediction can be put down as Eq. (2).

The P and B frame are derived from the content derived from I frame by the conditioning network. The input to the encoder is the GOP and the prediction process is supervised, the encoding used is binary motion encoding. The input to the encoder E(.) is the GOP is reduced by a factor 8 (height, weight and width). The output of the prediction is P&B frame in decoder is determined by the final output channel we set is encoder layer. The I0 frame determine the B frame and I_0 & I_{t+1} frame helps to predict the P frame. The loss in the process can be written as Eqs. (3) and (4).

$$
\begin{aligned}
\hat{B}_1 &= D[E(I_0, B_1, I_1) cond_0(I_0), cond_2(I_2)] \\
\hat{B}_2 &= D[E(I_0, B_2, I_2) cond_0(I_0), cond_3(I_3)] \\
\hat{B}_3 &= D[E(I_0, B_3, I_3) cond_0(I_0), cond_4(I_4)] \\
&\;\;\vdots \\
\hat{B}_{1,2,\ldots,t} &= D[E(I_0, B_{1,2,\ldots,t}, I_{t+1}) cond_0(I_0), cond_{t+1}(I_{t+1})] \\
&\forall x = 1, 2, \ldots, n
\end{aligned}
\tag{2}
$$

Here in Eq. (2), $\hat{B}_1, \hat{B}_2, \ldots, \hat{B}_x$ is predicted B frame or output, D is decoder, E represents encoder, $I_0, I_1, \ldots, I_t$ is the input in the $0^{th} 1^{st} \ldots . t^{th}$ frame that is the previous intraframe or I frame and current $B_1, B_2, \ldots, B_x$ represents the current bidirectional frame.

During training process, the I_0 is given directly but when the testing procedure comes the I_0 & I_{t+1} will be given from existing image codec.

6. Results and evaluation of proposed model

A main contribution is the collection of a dataset of currently more than 250 recordings (92 COVID-19, 73 bacterial pneumonia and 90 healthy controls). The dataset comprises data from various sources, including unpublished clinical data collected in a hospital by our collaborator in North Umbria (UK), as well as recordings from healthy volunteers scanned in Neuropil (Germany), but also data from other publications and educational websites. To avoid compression artifact the model is trained to or resized 64*64-pixel video frame. The train and validation split are like 435 and 40 clips respectively. The B net possess 18 frame length and P net possess 17. Both should predict 16 clips with a reconstruction loss expressed in Eqs. (3) and (4). The optimization used is Adam optimization. The epoch for training was 150, the training rate is 0.0001. the decayed factor is 2 at 30,100 & 140epoch. The loss of each training for prediction and bidirectional frame is represented in Eqs. (3) and (4).

$$L_R = \|B - \hat{B}\|^2 \tag{3}$$

$$L_R = \|P - \hat{P}\|^2 \tag{4}$$

The evaluation methods used are PSNR and SSIM. PSNR represents peak signal to noise ratio and SSIM with structural similarity index. The input of LUS dataset is evaluated

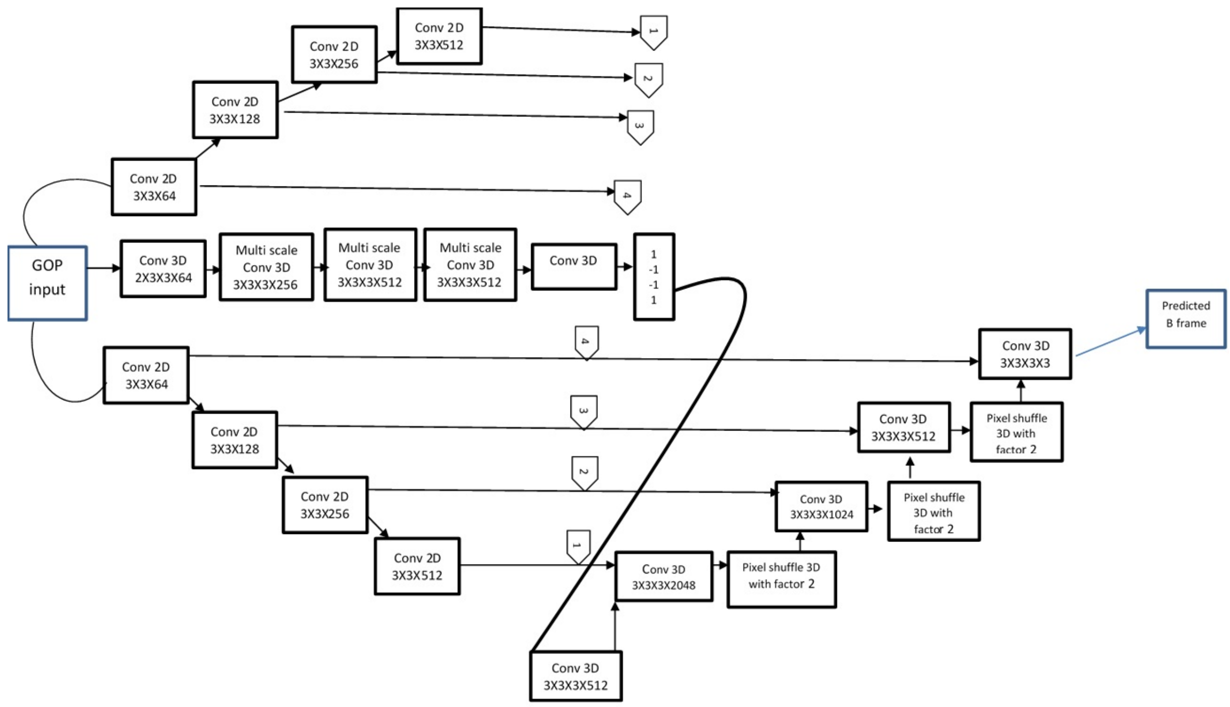

Fig. 6. The model of Deep CNN B-net.

with the existing HEVC (H.265), H.264 and our proposed model. The time taken for encoding its output etc are evaluated, this information is used to calculate the PSNR and SSIM. The PSNR is used to check the reconstruction quality of the video frame. It is calculated in db. The more the value of PSNR is the higher the similarity will be to the original video frame. SSIM does similar work as PSNR does, as it calculates the similarity index and its value lies between [-1,1].

The evaluation process is done by 253 raw video clips where each clip was partitioned in 17 frame/ 18 frame. The PSNR and SSIM of reconstructed video frames are calculated and the result obtained for HEVC, H.265 and proposed model is represented as graph in Fig. 7 for P-frame network. The graph clearly shows that the results are better for low bit rate or low bits per pixel for proposed model when compared to the H.265 and H.264.

The graph shows that at lower bit rates the design performs better. The graph shows that the B frame network can predict the bidirectional frame using B-net at lower bit rate better than the existing video codecs. The systems response is need to be improved for higher bit rates. The proposed model is thus great for low bit rate applications.

Fig. 8 shows that the B- net performs better for low to mid-range bit rate compared to H.265 and H.264. The results in the graph for B net shows that the bidirectional prediction using deep CNN shows a good mark for low and mid rage bit rates compared to exiting model for prediction and thus it can be replaced for the transfer of LUS data analysis and evaluation for COVID-19 and another lung disorders.

The proposed method P-Net and B-Net is compared with the existing techniques such as Exhaustive Search [20], Three Step Search [21], Diamond Search [22], Adaptive Rood Pattern Search [23] with the input dataset changed to macroblock of size 16x16 in Table 2 and 8x8 in Table 3. It clearly shows a better PSNR compared to the existing techniques. The applications [24] in the field are yet to explore.

7. Conclusion

Efficient video transfer and compression being an important area of research paved a way to work on the LUS data transfer through HEVC. The method proposes CNN based network to perform prediction by two designing two different network, the prediction frame network and bidirectional frame network. The inter prediction in HEVC compression is replaced by the CNN based P and B network. The network uses the GOP to predict P&B frame with multiscale CNN network. It replaces the block motion

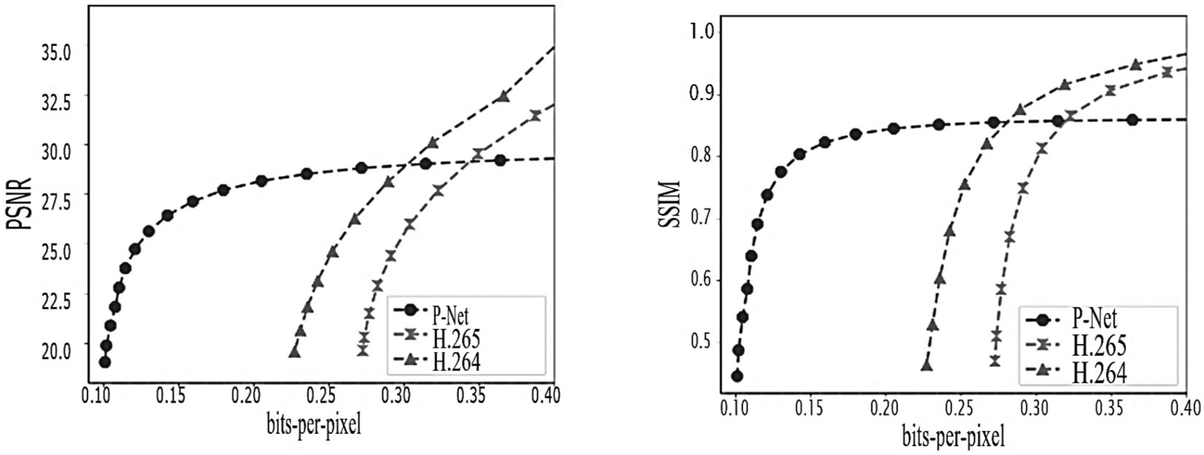

Fig. 7. Comparison graph of PSNR and SSIM for P-net with existing techniques.

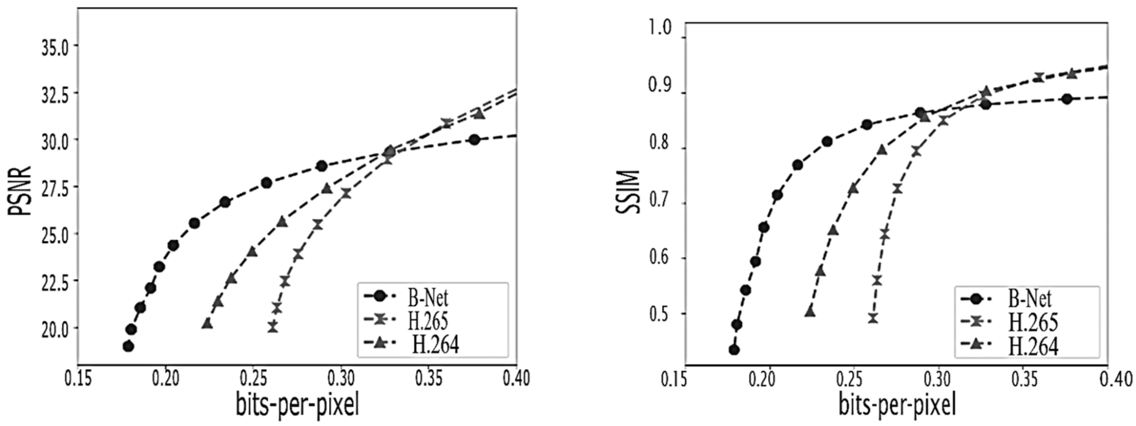

Fig. 8. Comparison graph of PSNR and SSIM for B-net with existing techniques.

Table 2. Motion compensation value for macroblock split16x16 with pixel around the block as 7 for16 frame video prediction.

Sl.no	Model	Bits per pixels	PSNR	SSIM	Time (sec)
1	Exhaustive Search [20]	0.0108	16.3	0.850	11.35
2	Three Step Search [21]	0.0108	16.3	0.851	1.53
3	Diamond Search [22]	0.0100	15.7	0.816	0.77
4	Adaptive Rood Pattern Search [23]	0.0097	15.7	0.816	0.63
5	P-Net	0.0052	28.9	0.829	0.28
6	B-Net	0.0038	30.3	0.859	0.28

Table 3. Motion compensation value for macroblock split 8X8 with pixel around the block as 7 for16 frame video prediction.

Sl.no	Model	Bits per pixels	PSNR	SSIM	Time (sec)
1	Exhaustive Search [20]	0.0582	19.4	0.901	46.32
2	Three Step Search [21]	0.0578	19.5	0.901	5.67
3	Diamond Search [22]	0.0552	18.6	0.860	2.82
4	Adaptive Rood Pattern Search [23]	0.0539	18.5	0.861	2.15
5	P-Net	0.0052	28.9	0.829	0.28
6	B-Net	0.0038	30.3	0.859	0.28

algorithm in existing video bit rate. The prediction process uses multiscale CNN network and conditioning network to predict P/B frame separately. B frame net requires $I_0 \& I_{t+1}$ while P frame net requires I_0 for prediction. The I frame is obtained from the existing image codec and the network explores its feature to obtain P/B frames. The results show that replacement of block based inter prediction algorithm with P-net and B-net will provide an advancement in inter prediction in lower bit rates than the existing methods. With these initial results, a major step is also to evaluate the approach on real clinical data. The model performs better for lower bitrates, it can be enhanced to work efficient in higher bitrates too.

References

[1] H. K. Joy and M. R. Kounte, (2020) *"A comprehensive review of traditional video processing"* **Advances in Science, Technology and Engineering Systems** *5*(6): 274–279. DOI: 10.25046/aj050633.

[2] S. Bouaafia, R. Khemiri, F. E. Sayadi, and M. Atri, (2020) *"Fast CU partition-based machine learning approach for reducing HEVC complexity"* **Journal of Real-Time Image Processing** *17*(1): 185–196. DOI: 10.1007/s11554-019-00936-0.

[3] S. Tanujaya, T. Chu, J.-H. Liu, and W.-H. Peng. "Semantic segmentation on compressed video using block motion compensation and guided inpainting". In: ***2020-October***. Cited by: 0. 2020.

[4] H. K. Joy, M. R. Kounte, and B. Sujatha, (2022) *"Design and Implementation of Deep Depth Decision Algorithm for Complexity Reduction in High Efficiency Video Coding (HEVC)"* **International Journal of Advanced Computer Science and Applications** *13*(1): 553–560. DOI: 10.14569/IJACSA.2022.0130168.

[5] G. Farnebäck, (2003) *"Two-frame motion estimation based on polynomial expansion"* **Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)** ***2749***: 363–370. DOI: 10.1007/3-540-45103-x_50.

[6] T. Shiodera, A. Tanizawa, and T. Chujoh. "Block based extra/inter-polating prediction for intra coding". In: ***6***. Cited by: 18. 2006, VI445–VI448. DOI: 10.1109/ICIP.2007.4379617.

[7] L. Zhao, S. Wang, X. Zhang, S. Wang, S. Ma, and W. Gao. "Enhanced ctu-level inter prediction with deep frame rate up-conversion for high efficiency video coding". In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. IEEE. 2018, 206–210.

[8] Y. Wang, X. Fan, C. Jia, D. Zhao, and W. Gao. "Neural Network Based Inter Prediction for HEVC". In: ***2018-July***. Cited by: 20. 2018. DOI: 10.1109/ICME.2018.8486600.

[9] N. Yan, D. Liu, H. Li, B. Li, L. Li, and F. Wu, (2019) *"Convolutional Neural Network-Based Fractional-Pixel Motion Compensation"* **IEEE Transactions on Circuits and Systems for Video Technology** *29*(3): 840–853. DOI: 10.1109/TCSVT.2018.2816932.

[10] S. Naveen, M. R. Kounte, and M. R. Ahmed, (2021) *"Low Latency Deep Learning Inference Model for Distributed Intelligent IoT Edge Clusters"* **IEEE Access** ***9***: 160607–160621. DOI: 10.1109/ACCESS.2021.3131396.

[11] J. Tian, F. Geng, F. Zhao, F. Gao, and X. Niu, (2022) *"Application of convolutional neural network in fault line selection of distribution network"* **Journal of Applied Science and Engineering (Taiwan)** *25*(1): 195–205. DOI: 10.6180/jase.202202_25(1).0020.

[12] S. Wang, L. Zhao, S. Wang, X. Zhang, S. Ma, and W. Gao, (2019) *"Enhanced motion-compensated video coding with deep virtual reference frame generation"* **IEEE Transactions on Image Processing** *28*(10): 4832–4844. DOI: 10.1109/TIP.2019.2913545.

[13] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, (2018) *"Reducing complexity of HEVC: A deep learning approach"* **IEEE Transactions on Image Processing** *27*(10): 5044–5059. DOI: 10.1109/TIP.2018.2847035.

[14] J.-K. Lee, N. Kim, S. Cho, and J.-W. Kang, (2020) *"Deep video prediction network-based inter-frame coding in HEVC"* **IEEE Access** ***8***: 95906–95917.

[15] M. M. Ho, J. Zhou, G. He, M. Li, and L. Li. "SR-CL-DMC: P-frame coding with super-resolution, color learning, and deep motion compensation". In: ***2020-June***. Cited by: 3. 2020, 538–542. DOI: 10.1109/CVPRW50498.2020.00070.

[16] W. Park and M. Kim, (2021) *"Deep Predictive Video Compression Using Mode-Selective Uni- And Bi-Directional Predictions Based on Multi-Frame Hypothesis"* **IEEE Access** ***9***: 72–85. DOI: 10.1109/ACCESS.2020.3046040.

[17] J. Lei, Z. Zhang, D. Liu, Y. Chen, and N. Ling. "Deep Virtual Reference Frame Generation for Multiview Video Coding". In: ***2020-October***. Cited by: 1. 2020, 1123–1127. DOI: 10.1109/ICIP40778.2020.9191112.

[18] J. R. Pare, I. Camelo, K. C. Mayo, M. M. Leo, J. N. Dugas, K. P. Nelson, W. E. Baker, F. Shareef, P. M. Mitchell, and E. M. Schechter-Perkins, (2020) *"Point-of-care lung ultrasound is more sensitive than chest radiograph for evaluation of COVID-19"* **Western Journal of Emergency Medicine** ***21***(4): 771–778. DOI: 10.5811/westjem.2020.5.47743.

[19] E. Shumilov, A. S. A. Hosseini, G. Petzold, H. Treiber, J. Lotz, V. Ellenrieder, S. Kunsch, and A. Neesse, (2020) *"Comparison of chest ultrasound and standard X-ray imaging in COVID-19 patients"* **Ultrasound International Open** ***6***(2): 36–40. DOI: 10.1055/a-1217-1603.

[20] I. E. Richardson. *The H.264 Advanced Video Compression Standard: Second Edition*. Cited by: 403. 2010. DOI: 10.1002/9780470989418.

[21] S. Kamble, N. Thakur, and P. Bajaj, (2017) *"Modified three-step search block matching motion estimation and weighted finite automata based fractal video compression"*:

[22] K. . .-. M. S. Zhu, (2000) *"Correction to "A New Diamond Search Algorithm for Fast Block-Matching Motion Estimation""* **IEEE Transactions on Image Processing** ***9***(3): 525. DOI: 10.1109/TIP.2000.826791.

[23] Y. Nie and K.-K. Ma, (2002) *"Adaptive rood pattern search for fast block-matching motion estimation"* **IEEE Transactions on Image Processing** ***11***(12): 1442–1449. DOI: 10.1109/TIP.2002.806251.

[24] S. Jha, S. Ahmad, H. A. Abdeljaber, A. Hamad, and M. B. Alazzam, (2022) *"A post covid machine learning approach in teaching and learning methodology to alleviate drawbacks of the e-whiteboards"* **Journal of Applied Science and Engineering (Taiwan)** ***25***(2): 285–294. DOI: 10.6180/jase.202204_25(2).0014.