HEVC intra prediction mode classification by deep learning

Hanxiang Wang*, Yanfen Li, L. Minh Dang, Hae Kwang Kim
Sejong University, Department of Computer Science and Engineering, Seoul, Korea

ABSTRACT

In High Efficiency Video Coding (HEVC) standard, the best intra prediction mode is decided by choosing the smallest rate-distortion cost of actual encoding among the total of 35 modes with the MPM (Most Probable Mode) scheme for compression purpose of mode encoding with reference to the adjacent reference blocks of the current prediction unit. This causes heavy computational complexity. In this paper, a deep neural network is conceived and experimented as a probable module for the intra prediction mode decision process inside of the HEVC encoding scheme. The neural network is trained and tested with a ground-truth dataset constructed from actual HEVC Intra encoding of original images. For the performance of the test, accuracy is used as the percentage of the correct mode output by the designed neural network to the ground-truth mode. The experimental results show that the neural network does not give good accuracy for the correct mode. However, accuracy increased when similar angle mode is considered as the correct mode. Also, the special modes of DC and Planar for MPM are analyzed in this paper.

Keywords: HEVC, intra mode, neural network, image coding

1. INTRODUCTION

In High Efficiency Video Coding (HEVC) standard, the best intra prediction mode is decided by choosing the smallest rate-distortion cost of actual encoding among the total of 35 modes with the MPM (Most Probable Mode) scheme for compression purpose of mode encoding with reference to the adjacent reference blocks of the current prediction unit. This causes heavy computational complexity [1]. There have been methods presented for fast HEVC intra prediction with traditional algorithmic ways [2] and with deep learning-based approaches [3][4]. These methods are integrated in the HEVC encoding scheme and showed its coding performance with experiments. For training and intra mode prediction of the current N x N size PU, the adjacent L lines of 2N pixels to the left and to the up of the PU are used together with the PU [3][4]. A full-connection layer-based model is used in [3] and a convolutional layer is used in [4]. As the contribution of this paper, the pure prediction performances of several neural network models are reported for HEVC intra prediction, which is not reported in the aforementioned neural network-based papers. The experimental information is expected to consider the plausibility of deep learning networks to be used in HEVC intra mode prediction process for reducing complexity. Section 2 explains in detail the work in this paper. Section 3 gives experimental results with some analysis. The paper concludes with section 4.

*1106452043@qq.com;

2. TESTED NEURAL NETWORKS

Figure 1 shows the *2* neural network models are used in this paper. The input layer is pixel data used for prediction and output

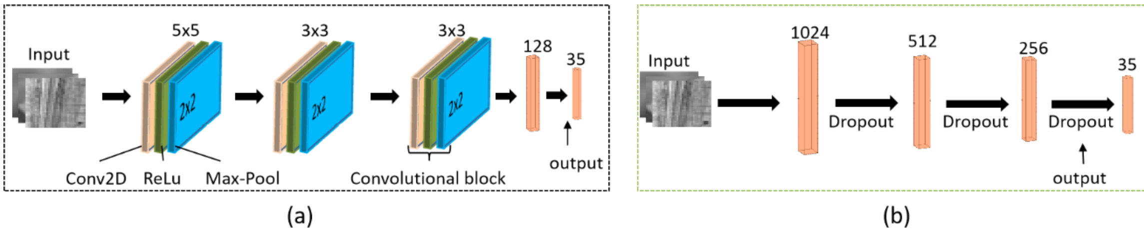

Figure 1. Two neural network models used in this HEVC intra mode prediction experiment: (a) Convolutional neural network (b) Fully connected neural network

layer is for 35 HEVC intra modes. Figure 1 (a) is a convolutional neural network-based model. Three convolutional layers are used in sequence with pooling followed by a final fully connected layer. (b) is a fully connected neural network.3 hidden layers are fully connected with dropouts. These are similar to the neural networks used in papers [3] and [4]. From a test set of images coded with HEVC, the prediction mode and its related pixels including the prediction unit are extracted as ground truth. Figure 2 (a) shows the input layer pixel data used. The intra mode is for the PU (prediction unit) of size *32*x*32*. *Top* is the horizontal pixel line of size *1*x*33* adjacent up to the PU. *Top-right* is the pixel line of size *1*x*32* to the right of *Top*. *Left* is the vertical pixel line of size *32*x*1* adjacent up to the PU. *Left-bottom* is the pixel line of size *32*x*1* to the bottom of *Left*. Only *PU, Top* and *Left* are used for the convolutional neural network. Figure 2 (b) shows the input layer pixel data for the fully connected network. This is obtained by flattening *PU*, *Top*, *Top-right*, *Left* and *Left-bottom* in (a) altogether as same as HEVC calculation of residual. The papers [3] and [4] uses more pixel information than HEVC with many top horizontal lines and left vertical lines.

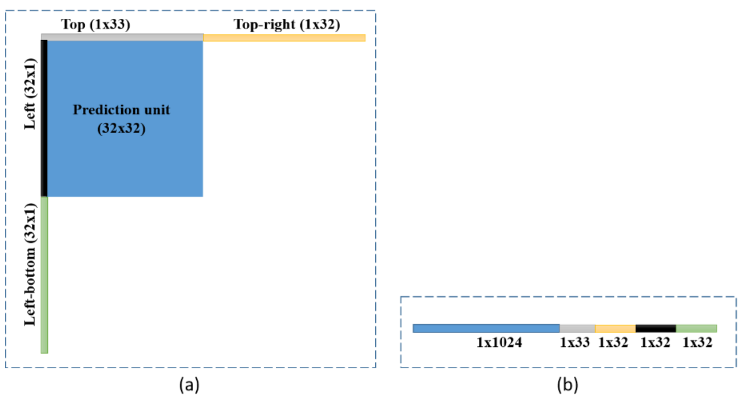

Figure 2. Input layer data for the neural networks. (a) is the input layer data in 2-dimensional layer for the convolutional network model (b) is the input layer data in 1-dimensional layer for the fully connected layer network model

3. EXPERIMENTAL RESULTS

To create a dataset, *500* images are randomly selected from ImageNet database [5]. These images are encoded with HM16.15 reference software. Only 32x32 size PU was aggregated. For the input layer pixel data, two groups of patches, one group with *12,373* PU only patches and the other group with *16,714* patches of PU together with adjacent pixel are extracted. Training, validation and testing dataset consists of 70%, 10% and 20% of the dataset. There are four experiments distinguished with input and out layers of CONV_ONE, CONV_MUL, CONV_EXT and FULL. For CONV_ONE, the input layer is PU-only patches and the output layer is softmax. For CONV_MUL, the input layer is PU-only patches and the output layer is sigmoid. For CONV_EXT, the input layer is PU with *Top* and *Left* adjacent pixel patches and the output layer is sigmoid. For FULL, the input layer is PU with *Top*, *Top-right*, *Left* and *Left-bottom* adjacent pixel patches and the output layer is sigmoid. The best intra mode is predicted from softmax output layer of CONV_ONE and top three best intra modes are predicted from sigmoid output layers of CONV_MUL, CONV_EXT and FULL. For CONV_MUL, CONV_EXT and FULL, if any of the *3* predicted intra modes are matched with the ground truth mode or its two neighboring angle modes, then it is counted as correct mode prediction. For example, if the best intra prediction mode is 23, then mode 24 and mode 22 are also correct modes except DC and Planar modes. The accuracy metric is the following:

$$Accuracy = \frac{the\ number\ of\ Correct\ mode\ prediction}{Total}$$

Table 1 shows the summary of performance evaluation for the four experiments. CONV_ONE use only PU pixel data as input and gives the worst performance for test data set, even though it achieves high accuracy for training dataset CONV_MUL gives enhancement over the accuracy due to tolerance on correctness. CONV_EXT enhances CONV_MUL with more input data with TOP and LEFT pixels. Best results come from FULL using the all pixel information used by HEVC intra mode coding.

Table 1 Accuracy results for 4 experiments.

Experiment	Training accuracy (%)	Testing accuracy (%)
CONV_ONE	80	26
CONV_MUL	75	63
CONV_EXT	76	67
FULL	83	74

Figure 3 shows the progress of accuracy and loss of the four experimental neural networks through epochs during the training phase. CONV_ONE shows that validation results worsen with the increase of epochs. From the analysis of experimental data, the prediction output modes are not exactly the same mode as the ground truth mode but still similar angle mode to the ground truth mode. The similar angle mode is meaningful for compression purpose. So, ground truth is extended to include the results with similar angle modes as explained for the accuracy metrics explained above. And for CONV_MUL, CONV_EXT and FULL, the models output the *3* best intra modes considering the HEVC intra mode mechanisms of the candidate list with *3* modes.

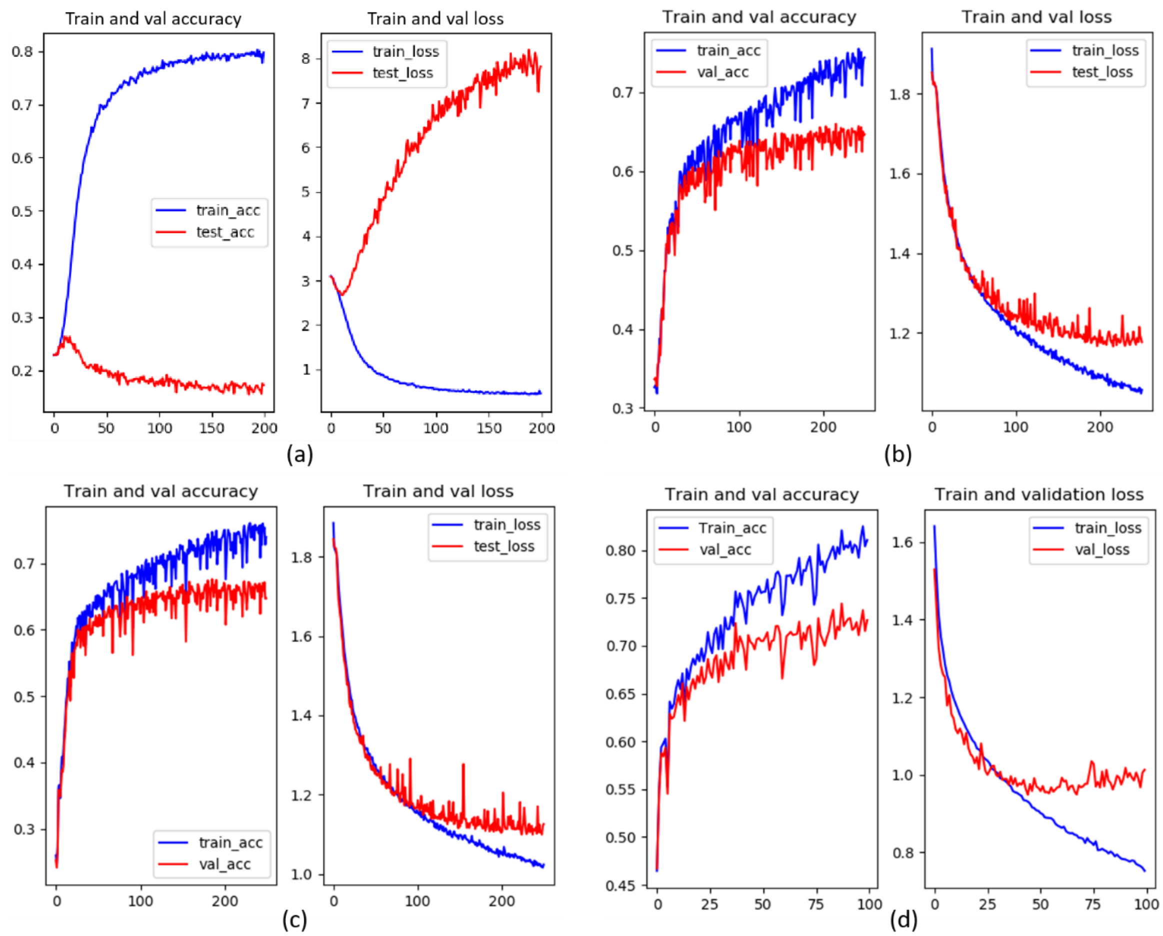

Figure 3. Accuracy and loss progress through epochs for training and validation of 4 experiments (a) CONV_ONE, (b) CONV_MUL, (c) CONV_EXT and (d) FULL

Figure 3 (d) shows that the accuracy increased gradually and reached approximately *80*% for training accuracy and at epoch 100^{th}, it achieves *70*% for validation accuracy, whereas the training loss decreases remarkably to under *0.8*. The validation loss has a sharp decrease in validation loss until epoch 45th. The results show that adding neighboring pixels improve the performance of the model. It's found that the majority of the modes produced by HEVC encoding are DC and Planar, so that it affects a lot on the performance evaluation and also to the real integration into the HEVC. So, it's very important to put efforts on the investigation of these two modes.

4. CONCLUSION

In this paper, four experimental neural models for HEVC intra mode prediction are introduced with its experimental results. It is found that not only PU data but also together with its neighboring pixel data contribute to better prediction results. Promising results are expected with more input data such as already decoded left and top neighboring intra mode used in HEVC intra coding. More intensive research is encouraged on DC and Planar modes due to their massive appearance in HEVC coded images. Research on reducing the network size is also a critical issue for actual integration into HEVC encoding pipelines.

ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1F1A1046236).

REFERENCES

[1] Sullivan, Gary J., Jens-Rainer Ohm, Woo-Jin Han, and Thomas Wiegand. "Overview of the high efficiency video coding (HEVC) standard." IEEE Transactions on circuits and systems for video technology 22, 1649-1668 (2012).

[2] Wang, Minghai, Hongan Wei, Ying Fang, Yanjie Song, and Zhongwen Zhuang. "Fast Mode Selection Algorithm for HEVC Intra Encoder." In 2018 IEEE International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), 169-172 (2018).

[3] Li, Jiahao, Bin Li, Jizheng Xu, Ruiqin Xiong, and Wen Gao. "Fully connected network-based intra prediction for image coding." IEEE Transactions on Image Processing 27, 3236-3247 (2018).

[4] Dumas, Thierry, Aline Roumy, and Christine Guillemot. "Context-Adaptive Neural Network-Based Prediction for Image Compression." IEEE Transactions on Image Processing 29, 679-693 (2019).

[5] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In 2009 IEEE conference on computer vision and pattern recognition, 248-255 (2009).