# Intra CTU depth decision for HEVC by using Neural Networks

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

**ABSTRACT**

As a video encoding standard, High Efficiency Video Coding (HEVC) achieves excellent performance while causing a dramatic increase in coding complexity. Especially, the coding tree unit (CTU) depth decision process is the most complicated section, which takes heavy computation complexity in the entire HEVC intra coding process. Therefore, a deep learning-based method is applied to directly predict the CTU depth level for each frame in this study. In addition, a large-scale dataset that contains the coding unit image files and the corresponding depths was generated by HM16.15 to train and test the deep learning model. Besides, a Convolutional Neural Network called LeNet is fine-tuned by modifying the original architecture, and then the model with a more complicated structure is evaluated and compared on an acquired dataset. The experiments show that the fine-tuned deep learning model has the ability to identify accurately the depth level of CTU, the recognition accuracy reaches over 98.6%.

**Keywords:** HEVC, Intra CTU, depth decision, neural network, image coding

## 1. INTRODUCTION

As a video encoding standard, High Efficiency Video Coding (HEVC) achieves excellent performance while causing a dramatic increase in coding complexity [1]. Especially, the coding tree unit (CTU) depth decision process is the most complicated section, which takes heavy computation complexity in the entire HEVC intra coding process. Recently, some deep learning-based methods have applied for CTU depth decision process. For example, the depth information of intra CTU is predicted by modifying a Convolutional Neural Network (CNN) model based on LeNet [2][5]. Similarly, a CNN model was designed to decide the current CU is split or non-split in [3]. Different with conventional CNN architecture, a novel structure with asymmetric kernels (AK-CNN) is adopted to classify the splitting decision of each CTU as well [4]. The main contributions of this paper are 1) the comprehensive performances of experimental models are reported in term of recognition accuracy and speed, which is not reported in the aforementioned studies. 2) Compared with [2], the fine-tuned models in this work can achieve higher accuracy. The experimental result is expected to introduce a CNN model that can be used to predict CTU depth in intra coding process for reducing computation complexity. Section 2 explains the detailed method used in this paper. The experimental results are showed and analyzed in section 3 to verify the performance of the presented method. The paper concludes with section 4.

* 1826535091@sju.ac.kr

## 2. INTRA CTU DEPTH DECISION METHOD

In this study, three experimental models are analyzed in terms of classification accuracy and speed. Fig.1 shows the neural network architecture of the optimized LeNet. In this paper, the input layer is pixel data used for prediction and output layer is 5 different classes described in Fig.2.

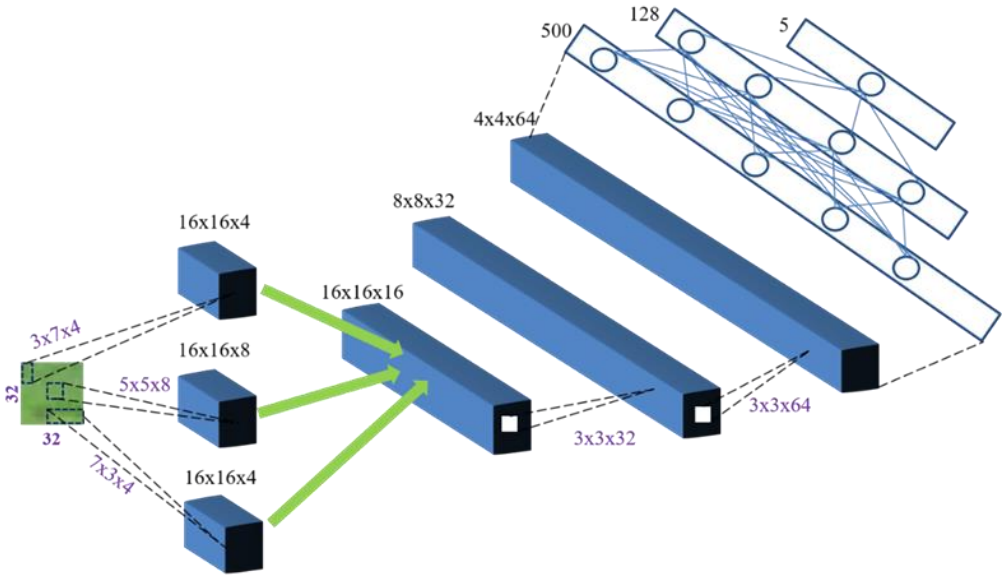


Figure 1. The neural network architecture of the optimized LeNet.

As proven in [4], the textures of near-horizontal and near-vertical are essential for the intra-prediction, which is also considerable for the partition result. Thus, it is significant to pay more attention to such texture patterns. For this reason, the proposed neural network extends the first convolutional layer to three different branches. The first branch is the traditional convolutional layer, with normal square kernels (5x5). In contrast, the remaining two branches have convolutional kernels of asymmetric shape that target at detecting near-vertical (3x7) and near-horizontal (7x3) textures. This structure could enable the neural network to identify the texture features more effectively and efficiently. The output of all three branches has an equal size. Next, we concatenate them and put them into two convolutional layers with small kernels to learn the correlation between these features. The combination of these features could help CNN better understand the content characteristics. Next, the extracted features flow through three fully connected layers to obtain the final prediction. In this neural network, the activation function of all convolutional layers and hidden fully connected layers is rectified linear unit (ReLU), while the activation function of the output layer is Softmax.

## 3. EXPERIMENTAL RESULTS

The dataset used in this study was generated from HEVC testing software HM16.15, including the 64x64 Coding Tree Units (CTU) image and depth information predicted by HEVC. The depth prediction represented by a 16x16 matrix. The elements in the matrix are 0, 1, 2 or 3, indicating depth 0, 1, 2, 3 for a 4x4 block in the CTU. After that, each CTU can be divided into 5 depth levels based on the depth matrix. As shown in Fig. 2, if the matrix contains only 0, then depth level is 1; if the matrix contains only 1, then depth level is 2; if the matrix contains 1 and 2, then depth level is 3; if the matrix contains 1, 2, and 3, then depth level is 4; if the matrix contains only 2 and 3, the depth level is 5. The strategy can enhance the encoding efficiency and retain more image details [6]. In this dataset, the number of images for 5 depth levels are 306,022, 191,458, 171,969, 249,857, and 488,963 respectively. During the training process, the ratio of training set to test set is 8:2.

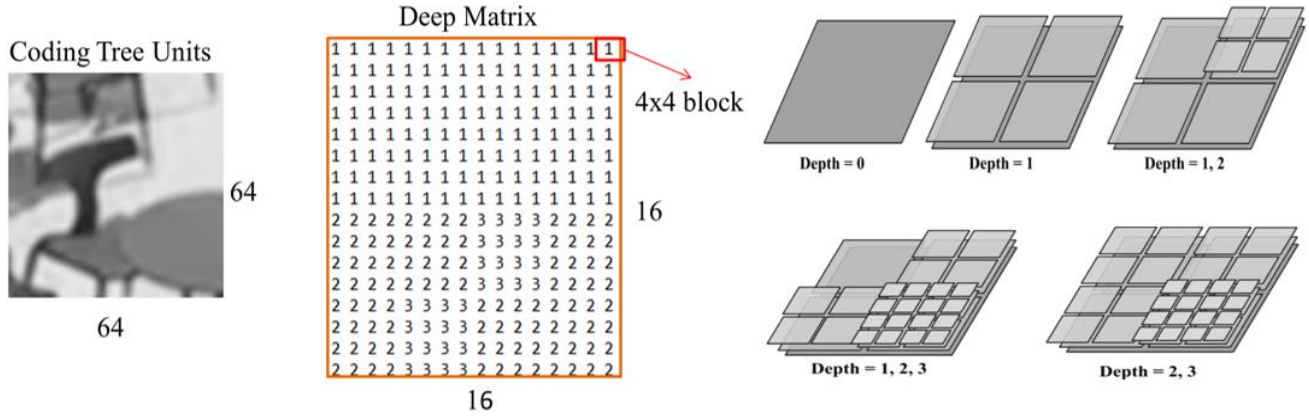


Figure. 2. The sample of 64x64 CTU image (left), depth information predicted by HEVC (middle), and the generated 5 classes data (right).

In this study, three distinct models are applied to evaluate the performances from the aspects of accuracy and speed. The light-weight model named LeNet is adopted and fine-tuned by using AK-CNN strategy on the acquired dataset. Apart from the original LeNet and the optimized LeNet, a classic CNN model VGGNet with a complicated structure is also tested and compared. As shown in Fig.3, the training performance, and validation performance are obtained by the original LeNet (a), VGGNet (b), and fine-tuned LeNet (c). In the training process of the original LeNet model, it occurs over-fitting problem at 28$^{th}$ epochs. However, the accuracy of the fined-tuned LeNet by using asymmetric kernels (AK-CNN) method increases steadily. The accuracy of fine-tuned model is 8% higher than that of the original model. Moreover, the classification performance of VGGNet based on the same dataset was presented in Fig.3 (c). It can be observed that the VGGNet achieved a better accuracy, which is 8.6% higher than the fine-tuned LeNet. Table 1 compares some information for LeNet and VGGNet. In terms of the computation complexity, the LeNet is lighter than VGGNet.

Table 1. Some computation information for original LeNet, VGGNet, and modified LeNet.

| Model | Total parameters | Trainable parameters | Time (for one CTU patch) |
|---|---|---|---|
| LeNet | 629，575 | 629，575 | 0.016s |
| VGGNet | 22, 146, 117 | 11, 560, 965 | 0.028s |
| Modified LeNet (Proposed) | 601，721 | 601，721 | 0.009s |

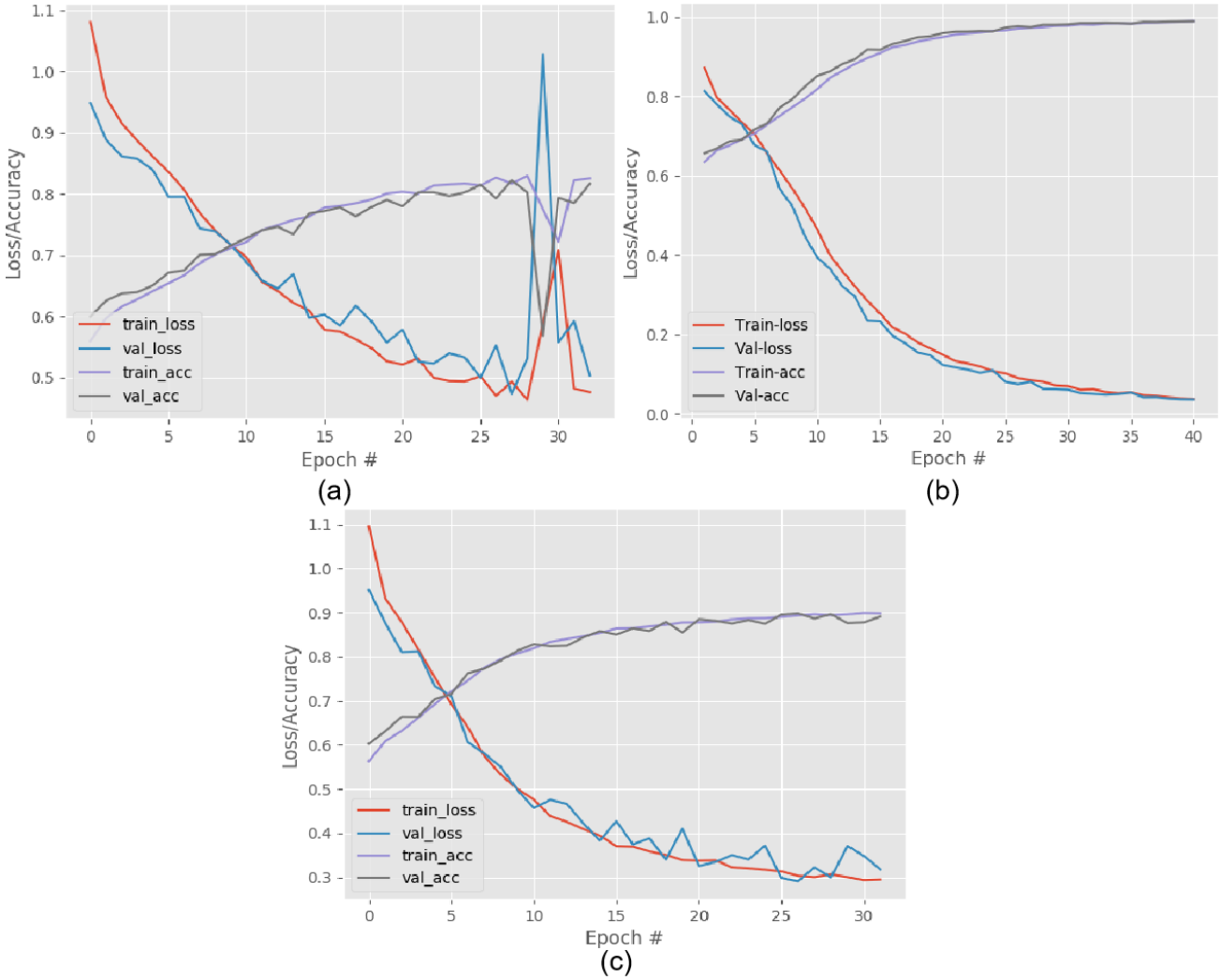


Figure. 3. Training loss, training accuracy, validation loss, and validation accuracy for original LeNet (a), VGGNet (b), and optimized LeNet (c).

## 4. CONCLUSION

In this paper, there distinct experimental neural models (LeNet, modified LeNet, and VGG-19) for CTU depth decision are introduced with the corresponding experimental results. It is observed that the highest accuracy was obtained by VGG-19, while the classification speed is slower than the other experimental models. In addition, the modified LeNet

performed better than the original LeNet model after using the asymmetric-kernel strategy. More intensive research is encouraged on the integration of the trained deep learning model and HEVC to evaluate the overall performance in HEVC coding process.

## 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] Feng, Zeqi, Pengyu Liu, Kebin Jia, and Kun Duan. "Fast intra CTU depth decision for HEVC." IEEE Access 6, 45262-45269 (2018).

[3] Kim, Kyungah, and Won Woo Ro. "Fast CU depth decision for HEVC using neural networks." IEEE Transactions on Circuits and Systems for Video Technology 29, no. 5, 1462-1473 (2018).

[4] Chen, Zhibo, Jun Shi, and Weiping Li. "Learned fast HEVC intra coding." IEEE Transactions on Image Processing 29, 5431-5446 (2020).

[5] Jia, Yangqing. "Training LeNet on MNIST with Caffe." (2019).

[6] Feng, Zeqi, et al. "Fast intra CTU depth decision for HEVC." IEEE Access 6, 45262-45269 (2018).