Grzegorz Pastuszak, Ph.D., D.Sc., university prof.
Institute of Radioelectronics and Multimedia Technology
Faculty of Electronics and Information Technology
Warsaw University of Technology
grzegorz.pastuszak@pw.edu.pl

16.06.2025

Reviewer's opinion
on Ph.D. dissertation authored by

Mateusz Lorkiewicz, M.Sc.,

entitled:

Block partitioning in video encoding with the use of artificial neural network

1. Problem and its impact

The subject of the review is the doctoral dissertation of Mr. Mateusz Lorkiewicz, M.Sc., entitled " Block partitioning in video encoding with the use of artificial neural network." The reviewed doctoral dissertation is experimental. In his doctoral dissertation, the Ph.D. candidate raised the problem of the encoding time reduction of the H.265/HEVC video encoder related to its partitioning algorithm. In particular, the author discussed the design of a partitioning algorithm using an artificial neural network (ANN) model and a decision algorithm. This scientific problem holds practical significance as it aims to keep encoding efficiency and provide control over the Encoding Time vs. Compression Efficiency trade-off. The author set two theses. The first concerns the benefits of using ANN to decrease computational complexity. The second states that the soft-decision algorithm applied to ANN results allows single-parameter control over encoding time and compression efficiency.

The objectives of the research, set in the doctoral dissertation, were formulated clearly and understandably. The problems raised are important and topical in the field of the problem, as they concern effective methods of controlling video encoders. Thus, the goals have been correctly defined and consistently meet the expectations of the scientific community dealing with video compression. They are ambitious and in line with current engineering and technical sciences challenges in the discipline of Technical Information Technology and Telecommunications.

2. Contribution

The dissertation is divided into 10 chapters. The author's contributions are described in Chapters 4-9. The main original contribution of the dissertation is the development of several variants of the partitioning algorithm that significantly reduces encoding time while maintaining almost the same compression efficiency. The partitioning algorithm is limited to software implementations for Intra modes. Although it was developed for H.265/HEVC, the general concepts can be bypassed in favor of the newer standard H.266/VVC.

The partitioning algorithm utilizes an ANN model and a non-trivial decision algorithm to optimize the encoding process. By focusing on the trade-off between Encoding Time vs Compression Efficiency, the research provides a solution for improving video encoder control algorithms. The research used the H.265/HEVC video coding standard and its software reference model HM version 16.23. Developing models for one of the latest standards points to the novelty of the described achievements. The reviewer agrees with the author regarding main and secondary contributions. The main and original contributions of the dissertation are as follows:

1. In Chapter 5, the author presents the main idea of the proposed partitioning algorithm. It discusses the overall approach, the problem definition for ANN training, the inputs and outputs of ANN, and the training procedure. Additionally, details of implementing the partitioning algorithm into the Modified HM are provided in Section 4.7.
2. Chapters 5 and 6 present two main approaches for partitioning algorithms: Basic and Extended. Detailed architectures of ANN are presented, along with training and evaluation results. The choice of hyperparameters is also discussed in these chapters. In the Basic architecture, the ANN estimates the most suitable quaternary-tree division of a 64x64 Coding Tree Unit (CTU) into Coding Units (CUs) based on luma samples included in the CTU. The extended architecture additionally predicts the division of 8x8 CUs into prediction units (PUs). The ANN generates the output tensor of probabilities for four (Basic) or five (Extended) division depth levels. The hard decision based on the maximal likelihood of CU depth shows a high efficiency of the ANN class prediction in terms of accuracy, recall, precision, and F1-score measures. Evaluations for test sequences reveal significant encoding time reduction (~56%) at small compression efficiency losses. The encoding time reduction is 56.08% and 70.49% for the Basic and Extended architecture, respectively. These reductions correspond to the average BD-Rate of 1.86% and 3.66%, respectively.
3. In Chapter 7, a detailed description of decision algorithms is presented. It defines hard-decisive variants and verifies the viability of soft-decisiveness. The hard decision is based on the sum of contributions from the Mx4x4 tensor produced by the ANN, where separate sums correspond to each M depth level. In the index-based variant (AlgIdx), contributions are zeros or ones, where one is assigned to the depth level with the maximal probability. In the probability-based variant (AlgPrb), contributions are probabilities assigned to each depth level. The soft-decisive approach involves indicating multiple partitioning patterns based on the output of the ANN. This approach allows for a more flexible decision-making process by considering a set of partitioning patterns instead of just one. By regulating the certainty of the ANN output with a single parameter, the soft-decisive approach aims to improve compression efficiency while controlling the trade-off between Encoding Time and Compression Efficiency. Proposed decision algorithms are evaluated, and control over the Encoding Time vs. Compression Efficiency trade-off is proposed. The AlgPrb approach slightly improves the compression efficiency of the hard decision without an impact on the complexity. In particular, BD-Rate is decreased to 1.80% and 3.44% for the Basic and Extended architecture, respectively. The soft-decisive approach allows further improvements in the compression efficiency at a slight increase or even decrease in the encoding time. In particular, BD-Rate decreased to 1.23% and 1.71%, while encoding time decreased by 1.3% and increased by 7.2% for the Basic and Extended architecture, respectively.
4. In Chapter 8, the best-proposed partitioning algorithms with state-of-the-art solutions are compared. It evaluates the effectiveness of the proposed algorithms in terms of encoding time reduction and efficiency compared to existing methods. The chapter provides insights into the performance of the proposed algorithms. Compared to state-of-the-art solutions, the proposed partitioning algorithms offer the best trade-off between encoding time reduction and coding efficiency. The results show that the proposed algorithms provide the most straightforward control over the trade-off between coding time and coding efficiency, highlighting their superiority in improving encoding efficiency and control over the trade-off.

The above contributions are practical as they can be applied in video encoders to significantly reduce encoding time at a relatively small deterioration in compression efficiency. The contributions were also published in the journal paper during the Ph.D. course.

Apart from the main contributions, the dissertation has less important ones:

1. In Subsection 3.4.2, the author proposed a novel metric for comparison of the partitioning algorithms that offer the control of the Encoding Time vs Compression Efficiency trade-off. These metrics are similar to Bjontegaard Delta metrics for Rate and PSNR. A relative decrease in encoding time is used instead of PSNR. The author's metrics are: ΔBD-RATE|TS and ΔTS|BD-Rate. The values of proposed metrics were calculated along with the graphical comparison of the methods with the trade-off control.
2. In Sections 5.4 and 6.4, the ANNs are tuned by exploring several architecture variants. In Section 5.4, tuning refers to adjusting the architecture of the Basic Approach model without affecting its complexity. This includes modifications like layer type adjustment, training data preparation, and learning rate optimization algorithm choice to enhance model performance without increasing complexity significantly. In Section 6.4, tuning involves making adjustments to the architecture of the Extended Approach model. This includes changes such as layer type adjustments, label smoothing, label weighting, and modifications to the complexity of Subnetwork A. These tuning efforts aim to improve model performance while considering the impact on model complexity and compression efficiency. Evaluation results show that most trials fail, and some provide slight improvement. The results indicate which studies should not be repeated.
3. Chapter 9 contains exploration experiments for the proposed partitioning algorithm. It aims to improve the performance of Basic and Extended approaches and presents minor research achievements. The chapter discusses enhancing the ANN, utilizing the contextuality of the encoding process, and proposing a method for global optimization of CTU partitioning using Hidden Markov Model (HMM). Evaluation results show that most trials fail, and some provide slight improvement. The results indicate which research directions should be continued or abandoned.
4. The proposed partitioning algorithm was evaluated using the author's modification of the HM reference software. The software allows the fast implementation of the partitioning algorithm, including ANN-based, without influencing the rest of the decision-making algorithms in the HM (Subsection 2.3.2). The software is available for other researchers to use.

3. Correctness

The doctoral dissertation is prepared correctly from the editorial point of view. The edition of the work is careful, i.e., the drawings showing the results of the analyses are clear. The same applies to the tables. The language of the dissertation (English) is correct. There are some minor typos, style errors, and uncorrected editorial remainders. These defects are not significant from the point of view of the substantive assessment of the work.

The author of the work correctly presented the starting point of the research, which is the origin, justification of the subject, and motivation. In general, the research methodology described in the dissertation is correct. The author presents a large number of results for various modifications he evaluated. However, some numerical inconsistencies and unjustified side claims are listed in Section 5 of the review.

4. Knowledge of the candidate

Chapters 2 and 3 are tutorial parts of the dissertation. The state of the knowledge is described in Chapter 2. The author delves into the High Efficiency Video Coding (HEVC) technology, focusing on partitioning the Coding Tree Units (CTU) in HEVC. It also discusses Rate-Distortion Optimization, providing a general description and insights into RD Optimization in the HEVC Test Model. This chapter is a foundational exploration of key aspects of video encoding and compression efficiency. The research methodology is given in Chapter 3, including assessing video encoder modification impact, preparing training datasets for ANNs, and evaluating modified encoders using test sequences. The chapter also discusses challenges, such as employing ANNs for partitioning methods and training them based on dedicated datasets.

The literature on problems addressed in the dissertation list is rich, and the author referred to many publications. In particular, the list of references includes over 230 items. Their number and importance seem sufficient to show the background of the work and the current state of the art.

The author clearly presents the logical sequence of the research cited in the work.

The tutorial chapters (2 and 3) and the numerous references confirm a general knowledge of the Ph.D. candidate in the field of video compression and artificial neural networks.

5. Other remarks[1]

1. In Chapter 1, with regard to standardization work, it was worth mentioning that, in addition to the MPEG organization, the ITU, which has its own designations for the H.26x standards, also participated in the work on the latest video compression standards.
2. Section 2.1: The claim that 95% of the coding time is partitioning is an exaggeration, considering that this is one of the dimensions that determine complexity. For example, a simplified prediction or transformation would also significantly reduce computational complexity by affecting it in other dimensions. Besides, the claim is not supported by evaluation results. In Tab. 7.2, no TS value achieves a reduction of 95%.
3. DM matrix – Division Matrix matrix – repetition
4. In Subnetwork B, the division into parts corresponding to 32x32 blocks was not justified by the results of the effectiveness evaluation. Why?
5. S_{cdl} is not specified.
6. The results of the Accuracy measure should be the sum of diagonal elements of the confusion matrix. However, they differ slightly, considering Tab. 5.1. and 6.1. Why?
7. Accuracy should match values in the last columns of Tab. 5.10, 5.11, 6.10, and 6.11.
8. The soft-decision algorithm in Subsection 7.3.2.1 should be specified in the order where points 3.1 and 3.2 are exchanged.
9. The noise formula in Section 9.5 (page 135) is not specified.
10. The use of "connected" on page 74 is an inappropriate word usage. More suitable is "related to".
11. The fifth column in Table 5.17 is described incorrectly: BD-Rate -> BD-PSNR.
12. Unfinished sentence on page 95: "... classification of bigger."
13. Missed zero in the BD-PSNR value on page 95. 0.4 dB -> 0.04 dB
14. Missed zero in the BD-Rate value on page 107: 0.12 dB -> 0.012 dB

15. In Fig. 9.5, the channel number of concatenated future maps (58) does not match the sum of concatenated channels (56).
16. The reference to Subsection 4.7.5 on page 70 is incorrect. It should be 4.7.3.
17. Section numbers (9.6 and 9.8) referred to at the beginning of Section 9.8 are wrong. They should be 9.5 and 9.6.
18. Wrong referred table numbers in the first paragraph of Section 8.3.
19. Unfinished sentence modification in the first paragraph of Section 9.3.

Suggestions:

1. HM partitioning is suboptimal. So, better results would be possible if all partitioning modes/sizes were evaluated to produce reference data like the bottom-up approach.
2. Decision algorithms in Subsections 7.2.1 and 7.2.2 are based on the maximal likelihood of the best CU depth levels. If the level increases, the likelihood at the lower depth is neglected. It seems that lower depths should be cumulated with the current one in a given block as they support the decision to use blocks without the splitting. Such a scenario should be verified. For example, the C variable in AlgIdx should be calculated with the condition $\leq$ "equal or smaller" instead of "equal." Similarly, the computation of S_L in AlgPrb should be modified to cumulate probabilities of lower levels with the current one. Furthermore, one can modify the algorithms to use different C/S_L values at each depth level. In AlgPrb, the values can cumulate probabilities related to lower levels. As a result, the decision on the splitting could result from comparing two joint likelihoods related to the current an lower levels on the one side and higher levels on the other.
3. The soft-decision algorithm seems to achieve better results when it is not constrained to consider successive depth levels. In particular, decision variables C and S_L would be compared between arbitrary depth levels.
4. Considering adjacent samples from neighboring blocks in the way applied in Fig. 9.5 is inefficient as it prevents the ANN from fully analyzing and combining the input context. The adjacent samples would be considered by extending the range of input samples to the left and top instead of the zero padding. Moreover, the extension would be to the top and bottom sides to consider the context of the following CTUs. Generally, better results would be achieved when all inputs are combined into a single tensor directed to Subnetwork A.
5. The ANN was trained for specific quantization parameters (QPs), so each QP requires a separate ANN parameter set. Extending the ANN to accepted QP as input would be more useful.

6. Conclusion

Taking into account what I have presented above and the requirements imposed by *Article 13* of *the Act of 14 March 2003 of the Polish Parliament on the Academic Degrees and the Academic Title* (with amendments)[2] and *Article 187 of the Act of 20 July 2018 – The Law on Higher Education and Science*, my evaluation of the dissertation according to the three basic criteria is the following:

[2] http://www.nauka.gov.pl/g2/oryginal/2013_05/b26ba540a5785d48bee41aec63403b2c.pdf

A. Does the dissertation present an original solution to a scientific problem? (the selected option is marked with **X**)

X	☐	☐	☐	☐
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

B. After reading the dissertation, would you agree that the candidate has general theoretical knowledge and understanding of the discipline of **Information and Communication Technology**, and particularly the area of multimedia technology?

X	☐	☐	☐	☐
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

C. Does the dissertation support the claim that the candidate is able to conduct scientific work?

X	☐	☐	☐	☐
Definitely YES	*Rather yes*	*Hard to say*	*Rather no*	*Definitely NO*

Grzegorz Pastuszak

Signature