Speeding up VP9 Intra Encoder with Hierarchical Deep Learning Based Partition Prediction

Somdyuti Paul, Andrey Norkin and Alan C. Bovik

The University of Texas at Austin
LIVE
Laboratory for Image & Video Engineering

NETFLIX

Outline

Introduction

- In VP9, 64×64 superblocks are partitioned recursively, possibly down to 4×4 blocks at four hierarchical levels.
- The rate-distortion optimization (RDO) based partition decision is a slow process owing to the combinatorial complexity of the partition search space.

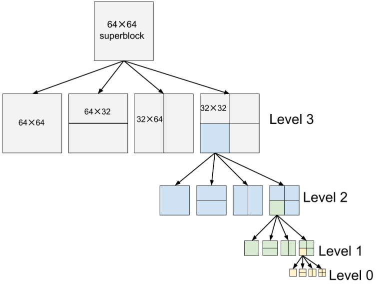

Figure 1: *Hierarchical superblock partition at four levels.*

Related Work

- Several machine learning (ML) based approaches with custom feature design attempted to reduce the computational overhead of the partition search in HEVC [1], VP9 [2] and VVC [3].
- Fewer works use deep learning based methods to solve the problem for HEVC [4, 5, 6].
- A parallel convolutional neural network architecture was employed in [4] to achieve a speedup of 61.8% for a 2.25% increase in BD-rate in the intra mode of HEVC.
- A multi stage ML-framework was used to sequentially make block partition decisions in [2], achieving a speedup of 60.1% over the speed 0 setting of the VP9 encoder with 0.07% increase in BD-rate.

Overview of Approach

Our approach involves a bottom-up block merge prediction using a hierarchical fully convolutional neural network (H-FCN) [7] .

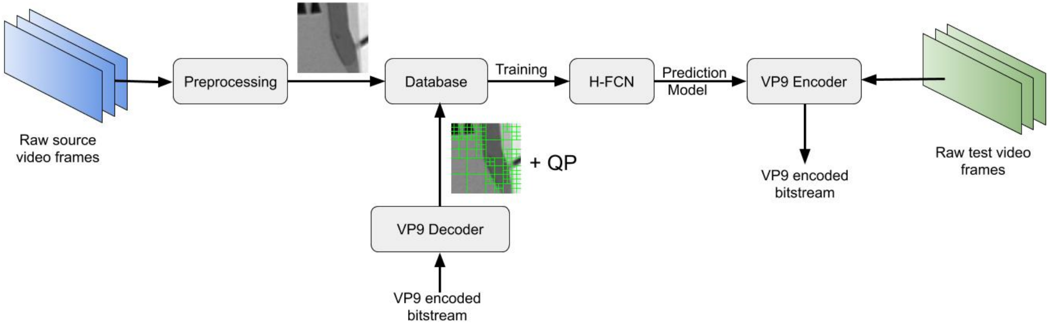

Figure 2: *VP9 partition prediction approach.*

implementation available at `https://github.com/Somdyuti2/H-FCN.git`

Database Creation

Content Selection

- The content for our database comprises 89 movies and 17 television episodes, which were selected from video sources in the Netflix catalog.
- Each video content was encoded at three different resolutions (1080p, 720p and 540p) using the reference VP9 encoder from the *libvpx* package.
- The contents were encoded in VP9 Profile 0, using speed level 1 and the *good* quality configuration.

- A concise description of the partition tree was required for effective learning.
- The partition tree was represented in the form of a set of four matrices:

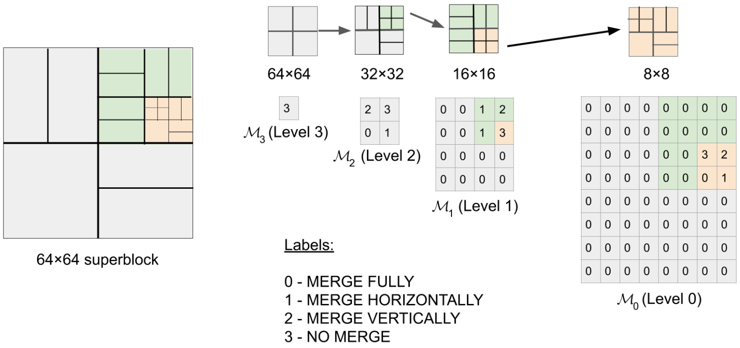

Figure 3: *Matrix representation of the four level partition tree.*

Database Creation

- The reference VP9 decoder from the *libvpx* package was modified to extract the superblock partition trees and the corresponding quantization parameter (QP) values from the encoded bitstreams.
- The raw pixel data for each superblock was obtained by extracting the luma channels of non-overlapping 64×64 blocks from the source videos downsampled to the encode resolution.
- Our database encompasses internal QP values in the range 8-105.

Table 1: *Summary of VP9 intra-mode superblock partition database*

Database	Contents	% of CGI content	# of samples
Training	62 (M) + 12 (E)	12.16	11 990 384
Validation	27 (M) + 5 (E)	12.50	4 698 195

H-FCN Model Architecture

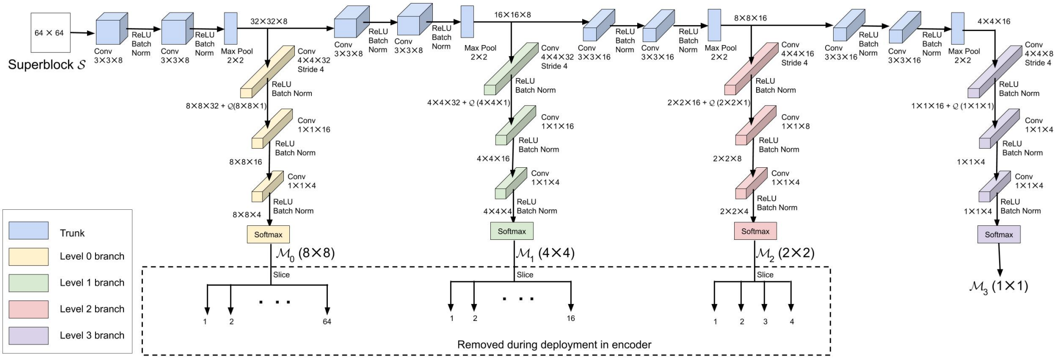

Figure 4: *Architecture of H-FCN model having 26 336 parameters and 54 610 FLOPs.*

H-FCN Training

Categorical cross entropy loss

$$L_q(\mathbf{w}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} y_{i,j}\log(p_{i,j}^{q}(\mathbf{w}))\ q = 1, \cdots, 85\ (N = 128, K = 4)$$

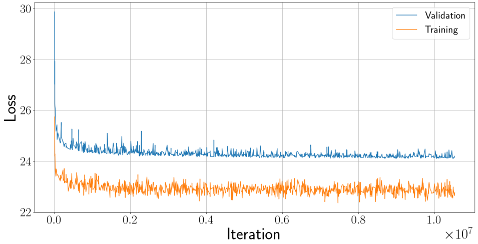

Figure 5: *H-FCN loss with training progress.*

Prediction Performance

The prediction accuracy at each level was evaluated on 10^5 randomly drawn samples from the training and validation sets.

Table 2: *Prediction accuracy of H-FCN model*

Level #	Training (%)	Validation (%)
0	89.42	90.27
1	84.42	83.47
2	86.07	85.13
3	91.73	91.18

Inconsistency Correction

- At each level, the model predictions are made independently of all other levels.
- Possible inconsistencies between the predictions of any two levels are corrected by a top-down approach.

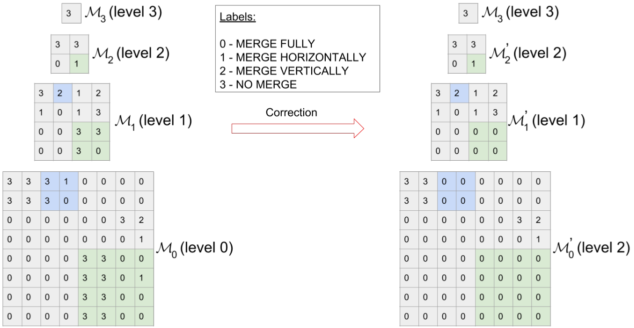

Figure 6: *Top-down inconsistency correction.*

Visualizing Superblock Partitions

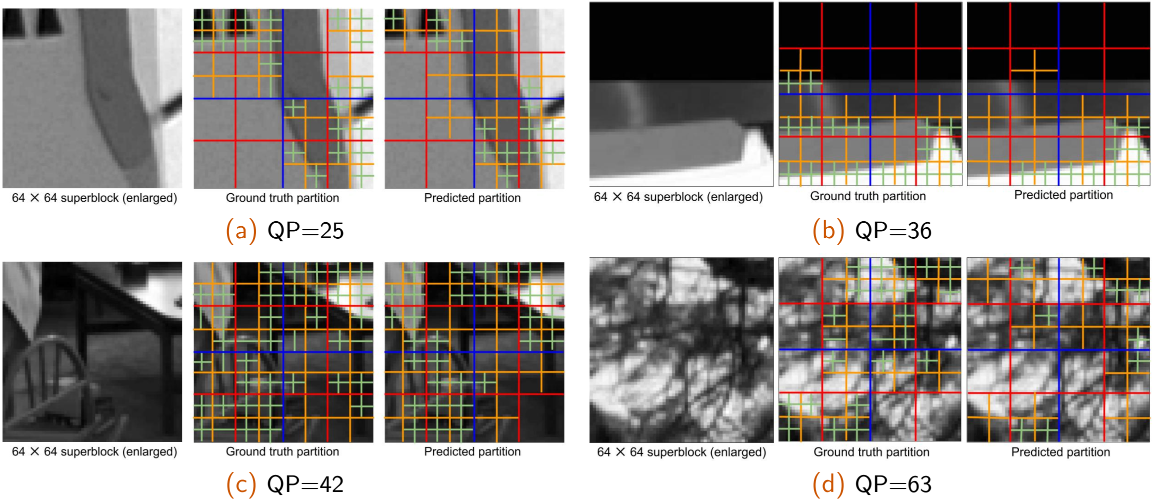

Figure 7: *Superblock partitions predicted by the trained H-FCN model compared with ground truth.*

Encoding Performance

- The trained model was integrated with the reference VP9 encoder using the Tensorflow C API.
- The predicted partitions were ordered to form a preorder traversal of the partition tree, and subsequently used to replace the RDO based partition decision in a recursive fashion.
- The encoding performance was evaluated on 30 test sequences at 3 resolutions in terms of both BD-rate and speedup (ΔT).

Table 3: *Encoding perfomance with respect to RDO baseline*

Resolution	ΔT (%)	BD-rate (%)
1080p	67.5	1.70
720p	72.2	1.75
540p	69.5	1.68
Overall	69.7	1.71

Encoding Performance

Comparison with Speed Level 4 of Reference Encoder

- The speedup and BD-rate of our approach was also compared with speed level 4 of the reference VP9 encoder, the highest recommended speed level for the baseline configuration.

Table 4: *Comparison of speedup versus BD-rate tradeoff of our approach with VP9 speed level 4*

Resolution	ΔT (%)		BD-rate (%)	
	Speed 4	H-FCN	Speed 4	H-FCN
1080p	62.0	67.5	2.95	1.70
720p	68.2	72.2	4.12	1.75
540p	65.9	69.5	2.38	1.69
Overall	65.4	69.7	3.15	1.71

Encoding Performance

Comparison with Speed Level 4 of Reference Encoder

The benefit offered by our approach in terms of speedup persists across the range of QP values used to learn the H-FCN model.

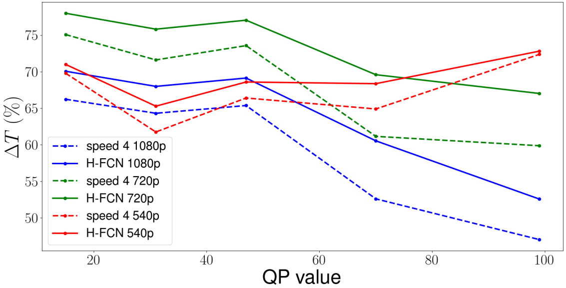

Figure 8: *Speedup achieved by H-FCN and RDO at speed 4 relative to baseline.*

Concluding Remarks

- Our H-FCN based partition prediction approach achieved 69.7% speedup on average at the expense of 1.71% increase in BD-rate.
- It achieves 4.3% higher speed up than the speed level 4 of the reference encoder, while incurring 1.44% smaller BD-rate penalty.
- Further benefits can possibly be derived by extending the approach to the AV1 codec.

References

[1] D. Ruiz-Coll, V. Adzic, G. Fernandez-Escribano, H. Kalva, J. Martinez, and P. Cuenca, "Fast partitioning algorithm for HEVC intra frame coding using machine learning," in *Proc. IEEE Int. Conf. Image Process.*, pp. 4112–4116, 2014.

[2] H. Su, C. Tsai, Y. Wang, and Y. Xu, "Machine learning accelerated partition search for video encoding," in *Proc. IEEE Int. Conf. Image Process.*, pp. 2661–2665, 2019.

[3] T. Amestoy, A. Mercat, W. Hamidouche, D. Menard, and C. Bergeron, "Tunable VVC frame partitioning based on lightweight machine learning," *IEEE Trans. Image Process.*, 2019.

[4] M. Xu, T. Li, Z. Wang, X. Deng, R. Yang, and Z. Guan, "Reducing complexity of HEVC: A deep learning approach," *IEEE Trans. Image Process.*, vol. 27, pp. 5044–5059, Oct. 2018.

References

[5] Z. Liu, X. Yu, Y. Gao, S. Chen, X. Ji, and D. Wang, "CU partition mode decision for HEVC hardwired intra encoder using convolution neural network," *IEEE Trans. Image Process.*, vol. 25, pp. 5088–5103, Nov. 2016.

[6] K. Kim and W. Ro, "Fast CU depth decision for HEVC using neural networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, pp. 1462–1473, May 2018.

[7] S. Paul, A. Norkin, and A. Bovik, "Speeding up VP9 intra encoder with hierarchical deep learning based partition prediction," *arXiv preprint arXiv:1906.06476*, 2019.