Convolutional Neural Network based Fast Intra Mode Prediction for H.266/FVC Video Coding

Ting-Lan Lin[1], Kai-Wen Liang[2], Jing-Ya Huang[2], Yu-Liang Tu[2], and Pao-Chi Chang[2]
[1]*Department of Electronic Engineering, National Taipei University of Technology, Taipei* , Taiwan
2Department of Communication Engineering, National Central University, Taoyuan , Taiwan

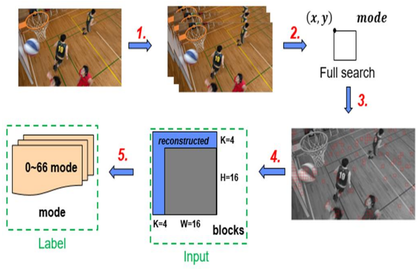
Abstract

The next-generation video compression standard H.266/Future Video Coding (FVC) provides high compression efficiency in terms of the cost of computing the optimal intra mode from 67 modes. We propose an intra mode prediction method based on a convolutional neural network (CNN). An input image set of 20 × 20 blocks is used to train the CNN; the CNN is used to predict the best classes of intra mode direction. The CNN architecture comprises two convolutional layers and a fully connected layer. Compared with the default fast search method in FVC, the proposed method can achieve a 0.033% decrease in Bjøntegaard delta bit rate (BDBR)with only a slight increase in time.

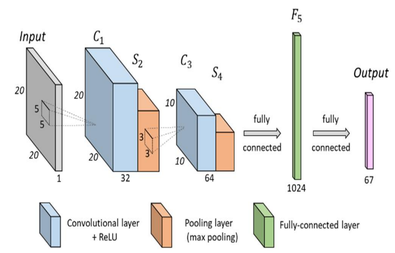
1. Introduction

➢ Data preprocessing and CNN model

1.1 Train Data Preprocessing

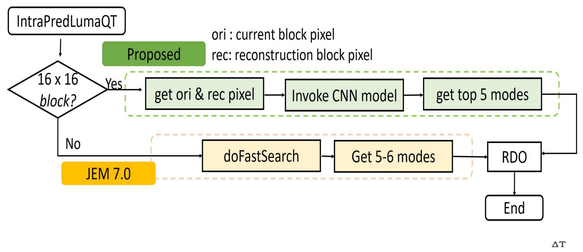


1.2 CNN Training Model



2. Proposed Method

We propose an intra mode prediction method based on a convolutional neural network (CNN). The CNN architecture comprises two convolutional layers and a fully connected layer. Only 16 × 16 blocks with the deep learning methodology were tested by JEM 7.0 in the simulations. Eighteen video sequences categorized into five classes (A to E) were tested using the Intra_main_10 encoding configuration, and each sequence was encoded 4 times with different quantization parameters (QP) equal to 22, 27, 32, and 37.



Overview flowchart of proposed method.

3.Exerimental Results

- *JEM 7.0 is used in the simulations.*
- *Eighteen video sequences categorized into five classes (A to E) are tested using the Intra_main_10 encoding configuration.*
- *Batch size is set to 2000 (blocks) with 20,000 epochs and a learning rate of 0.001. In testing, frames excluded from the training set are chosen to be encoded.*
- *QP : 22, 27, 32, and 37*

CNN models training	
CPU	Intel Xeon E5-2630v3 @2.4GHz * 2
GPU	GTX TITAN X
RAM	256GB DDR4-2133 MHz
OS	Ubuntu-x64
Software language	Python2.7、Tensorflow
Data feature extraction tool	Matlab2015a
Video coding	
CPU	Intel(R) Core(TM) i7-6800 @3.40 GHz
RAM	32.0 G bytes
Reference ΔT software	JEM 7.0 (C/C++)
Video coding tool	Visual Studio 2013 , Tensorflow , Python

Class	Sequence	doFastSearch in JEM7.0			Proposed method		
		BD-PSNR	BDBR(%)	(%)	BD-PSNR	BDBR(%)	(%)
A1 (4096x2160)	ToddlerFountain	-0.044	0.841	-94.67	-0.064	1.208	-93.21
A2	TrafficFlow (3840x2160)	-0.025	1.361	-94.14	-0.031	1.614	-92.64
	Rollercoaster (4096x2160)	-0.014	1.071	-94.46	-0.050	1.298	-98.28
B (1920x1080)	Kimono	-0.028	0.907	-94.46	-0.028	0.896	-94.26
	ParkScene	-0.046	1.046	-93.87	-0.010	0.929	-93.40
	BasketballDrive	-0.022	1.141	-94.47	-0.022	1.042	-94.11
	BQTerrace	-0.056	0.900	-93.66	-0.053	0.852	-93.35
C (832x480)	BasketballDrill	-0.108	2.301	-93.96	-0.103	2.165	-93.68
	BQMall	-0.065	1.053	-93.64	-0.063	1.025	-93.66
	PartyScene	-0.095	1.169	-92.56	-0.088	1.136	-92.21
	RaceHorsesC	-0.075	1.142	-93.28	-0.076	1.142	-93.14
D (416x240)	BasketballPass	-0.080	1.234	-93.93	-0.056	0.928	-93.64
	BQSquare	-0.132	1.517	-92.65	-0.126	1.484	-92.58
	BlowingBubbles	-0.104	1.700	-93.11	-0.090	1.469	-92.28
	RaceHorses	-0.119	1.611	-93.24	-0.106	1.428	-93.16
E (1280x720)	FourPeople	-0.076	1.319	-94.27	-0.071	1.232	-93.98
	Johnny	-0.051	1.323	-94.24	-0.049	1.244	-93.93
	KristenAndSara	-0.077	1.536	-94.15	-0.073	1.474	-93.92

4. Conclusions

We propose a CNN-based intra mode prediction method for H.266/FVC. The proposed method achieves a 0.033% BDBR decrease compared with doFastSearch in JEM 7.0. Although the time saving in JEM is not as great as that of doFastSearch, the method still saves over 93% of encoding time. The proposed method provides an alternative approach for expediting intra mode selection in H.266/FVC. This study test only 16 × 16 blocks with the deep learning methodology. More remarkable results for blocks of different sizes are expected through the application of appropriate neural network architectures Future studies should extend CNN structures to other parts of the coding process that require long search times for prediction, such as inter prediction and motion estimation.