A Machine Learning Approach to Accurate Sequence-Level Rate Control Scheme for Video Coding

Yangfan Sun, Mouqing Jin, Li Li and Zhu Li

Multimedia Computing & Communication Lab

Dept of Computer Science & Electrical Engineering

University of Missouri, Kansas City

Outline

- Motivation
- Introduction
- Overall scheme of proposed framework
- Content-dependent model
- Machine learning based model parameter fitting
- Experimental results
- Conclusion and future work

Background

- Constant rate factor (CRF) can provide constant quality during encoding process. It compresses different frames by providing varying QP due to taking the motion into account.
- Werner et al compare CRF and constant QB (CQB) given the conclusion is a certain improvement on saving bitrate by using CRF.

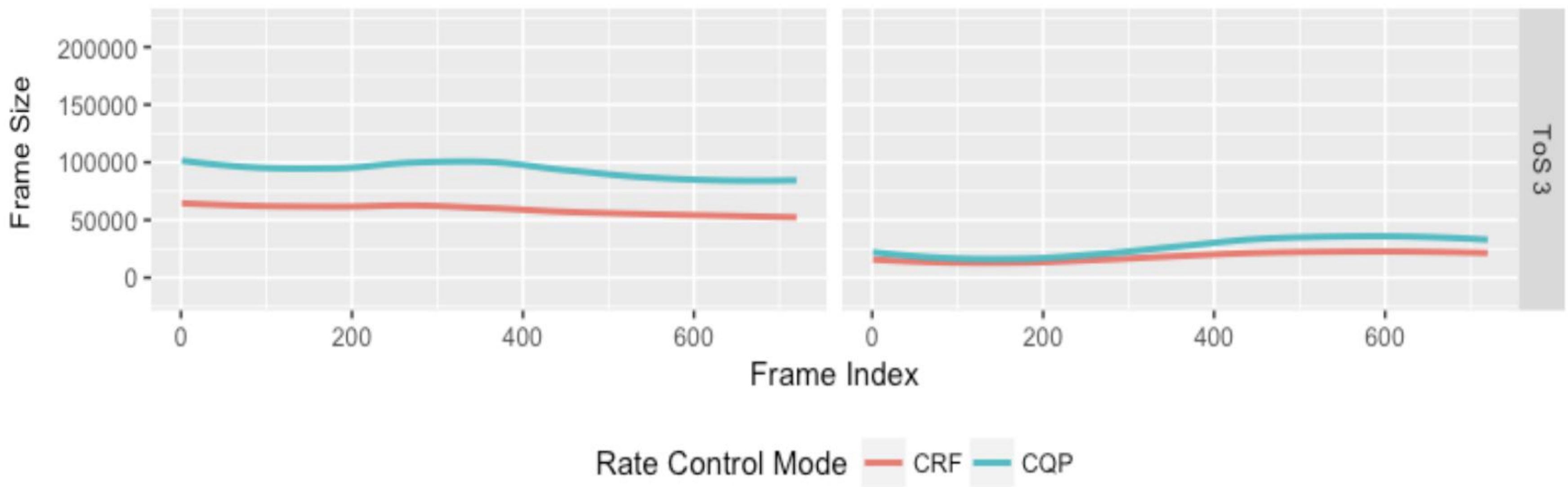

Fig 1. Comparation of CRF (red lines) and CQP (blue lines) at different level (17 and 23).

Motivation & Introduction

- Since better performance of CRF, we decide to employ CRF instead of QP as the parameter for rate control.
- However, the there is no obvious relationship between CRF and bitrate compared with the QP. A possible method is using multi-pass encoding which may lead to much higher complexity burden.
- In this paper, the main contributions as followed:
  - We found a robust content-dependent relationship between bitrate and CRF.
  - Then we applied neural network algorithm to avoid multi-pass encoding for approaching target bitrate.

CRF-Rate Model for Better Transcoding

- We intent to obtain the target bitrate by two-pass coding:
  - In 1st pass coding & offline training process: Deriving the model by using the coding features as the input of neural network to train content-dependent parameters for model.
  - In 2nd pass coding: With the assistant of derived model, we can infer bitrate after encoding target samples.

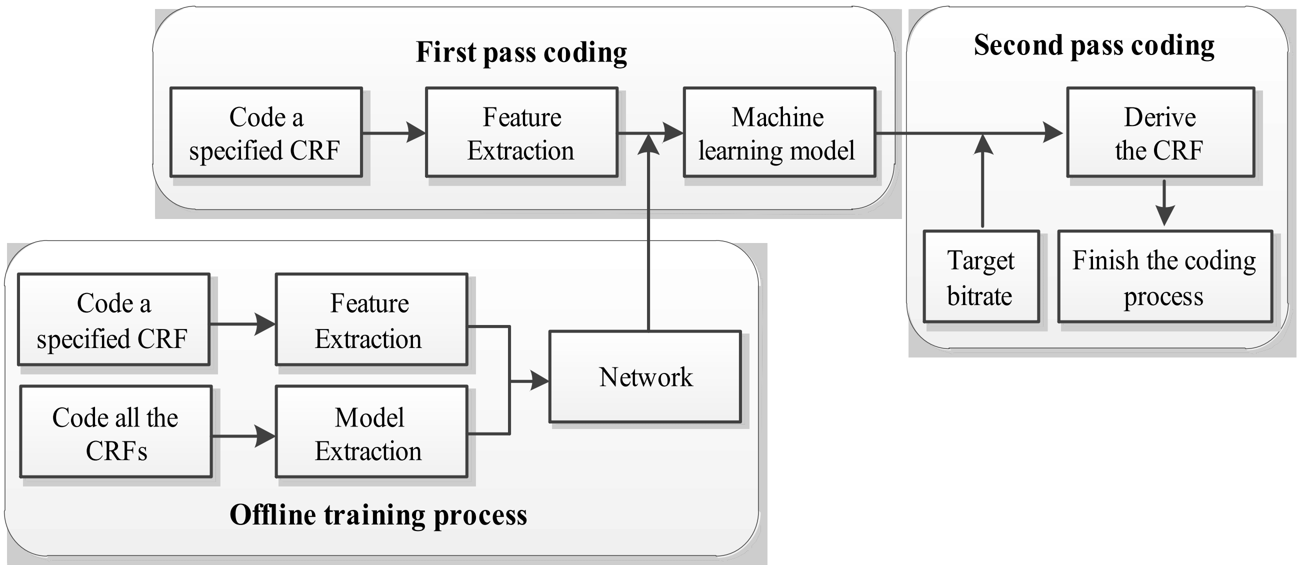

Fig 2. Framework of proposed scheme

Our Proposed CRF-R model

- We established a CRF-R model to depict the relationship between CRF and bitrate. Previous research presented a linear CRF-R model by Google, in our project, we presented 2nd order CRF-R model instead.
- Both model functions list below. $a(v)$, $b(v)$ and $c(v)$ are considered as content-dependent parameters,
  - Linear model:

$$CRF = a(v) \cdot \ln R + b(v)$$

  - 2nd order model (We proposed):

$$CRF = a(v) \cdot (\ln R)^2 + b(v) \cdot \ln R + c(v)$$

- We evaluated curve fitting between proposed model and linear model by least square method. The curves of each model and ground truth are shown on Fig. 2. The curve linear model obviously fits worse than our proposed model.

Improvements on proposed CRF-R model

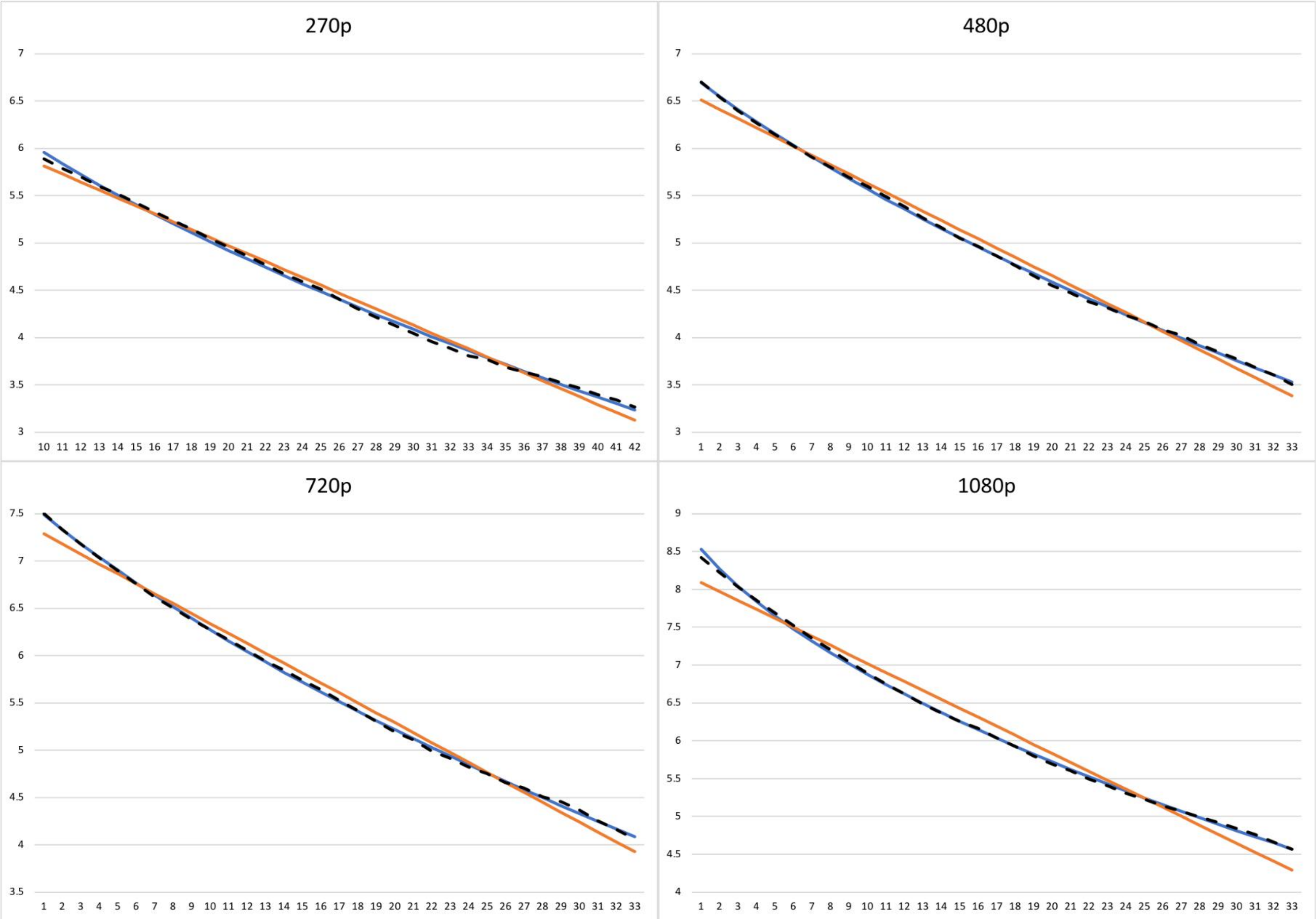

Fig. 2. The curve of linear model (orange lines) and our proposed model (blue lines) vs. ground true (black dashed lines), X axis represents CRF and Y axis represents ln(bitrate).

Improvements on proposed CRF-R model

- Table 1: shows the gains on bitrate estimative error we simply obtained from proposed CRF-R model.

	2nd order	Linear	2nd order	Linear
	Bitrate error within 20%		Bitrate error within 10%	
270p	100%	98% (2%)	99%	87% (12%)
480p	100%	96% (4%)	99%	80% (19%)
720p	99%	89% (10%)	97%	62% (35%)
1080p	97%	82% (15%)	89%	49% (40%)
Avg	99%	91% (8%)	96%	69% (27%)

Table 1. The ratio of relative bitrate error on each model, red numbers represent the improvement.

Content-dependent features introduction

- The bitrate allocation at sequence level depends on the complexity of content and information redundancy. Therefore, we extracted features involving information from both factors, shown as Fig. 3.

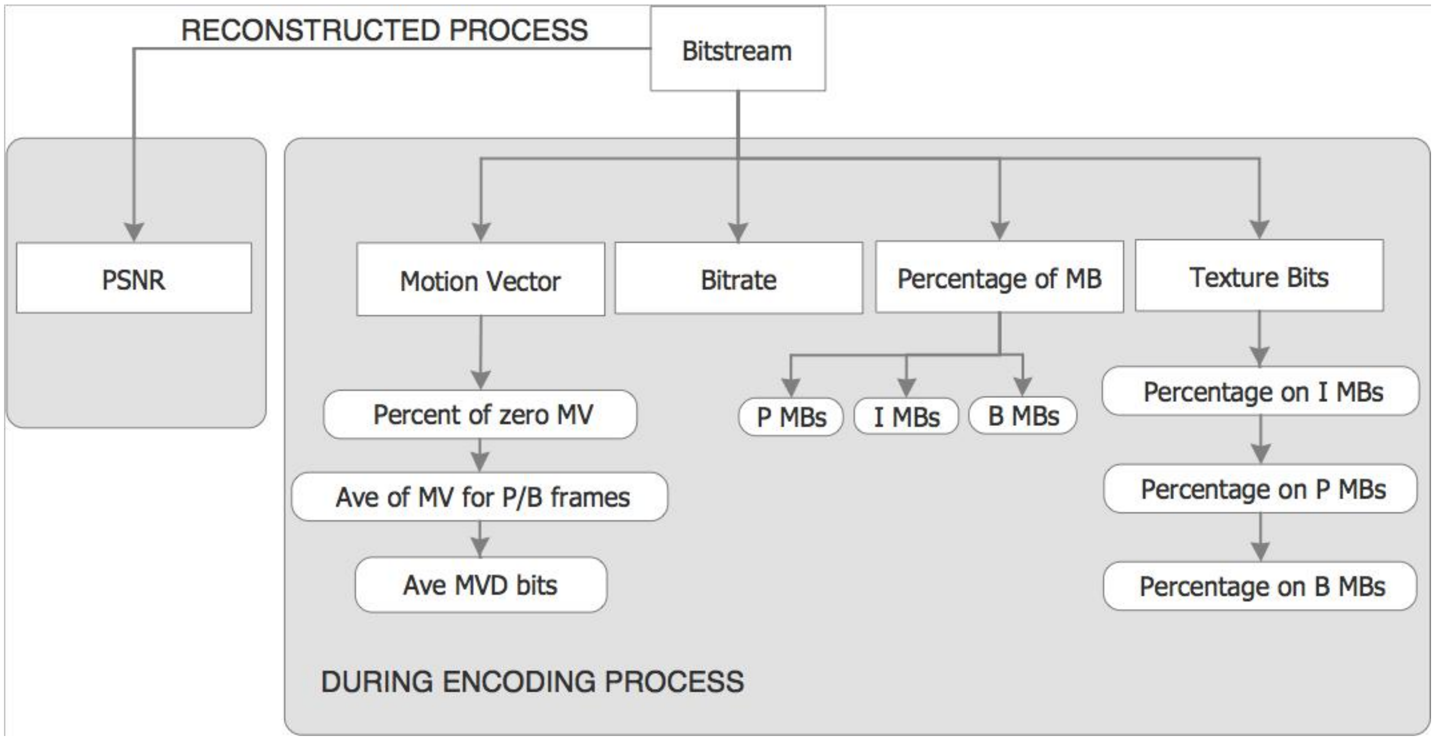

Fig. 3. 14 Features extracted for machine learning methods rate control

Neural network

- In this project, we adopt 5-sec video segmentation with various complexity levels of content at spatial and temporal domain to extract feature and feed in our shallow neural network.

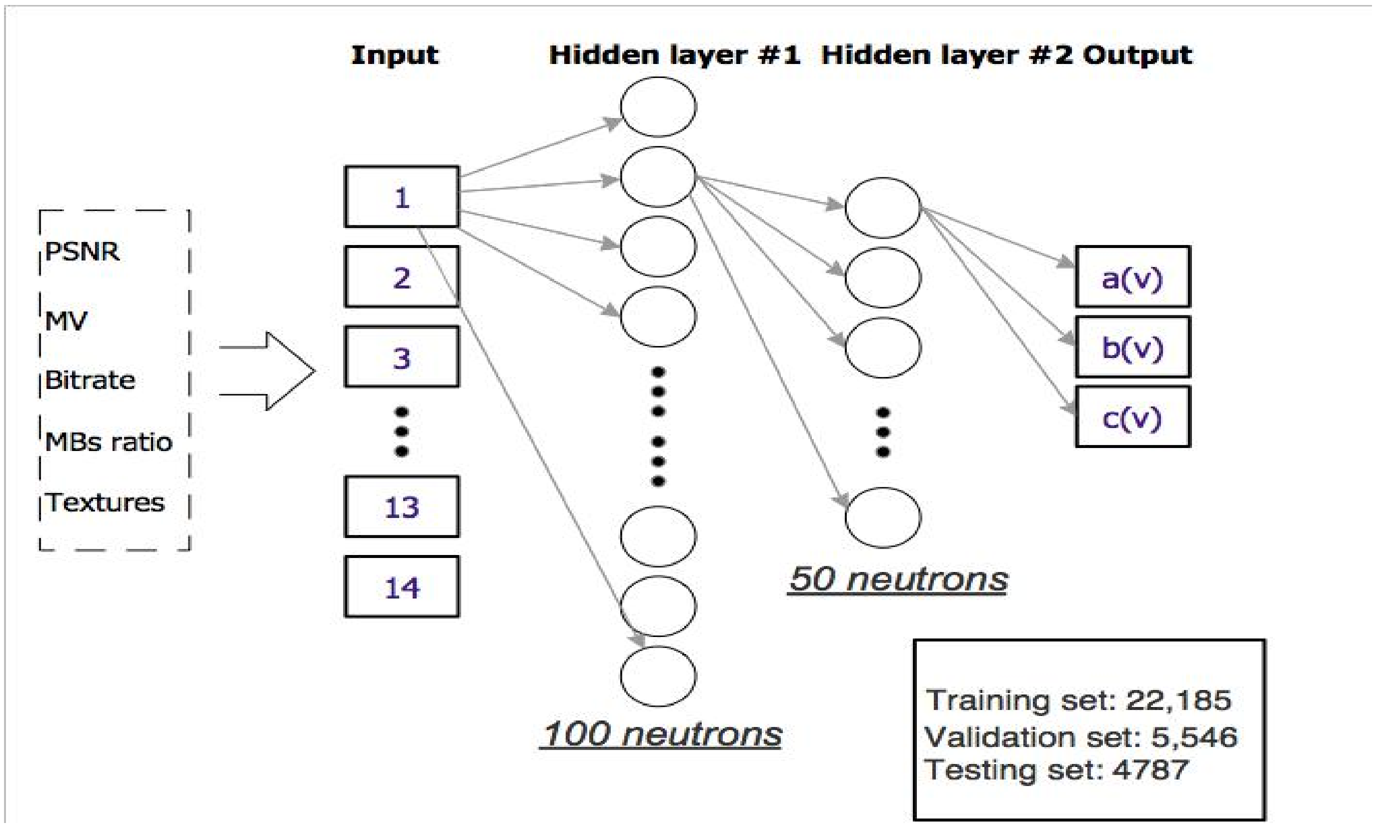

Fig. 4. the structure of proposed shallow neural network

Experimental results

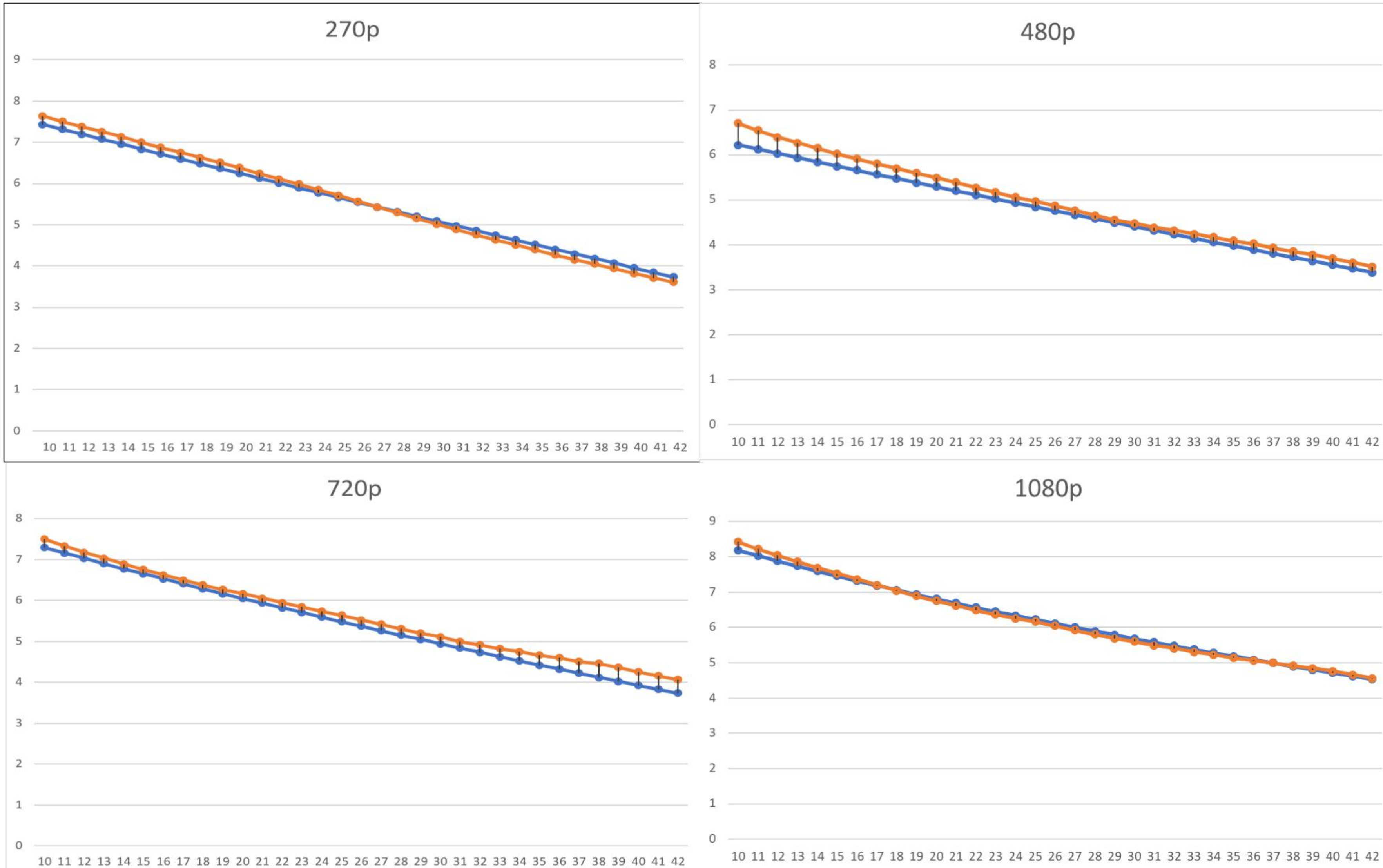

Fig. 2. The curve of predicted by SNN (orange line) vs. ground truth (blue lines), X axis represents CRF and Y axis represents ln(bitrate).

Experimental results

- We adopt massive iteration in order to guarantee the robustness of our method, training samples at each resolution respectively.

	2nd order	Linear	2nd order	Linear
	Bitrate error within 20%		Bitrate error within 10%	
270p	93.5%	95.0%(-1.5%)	72.9%	70.0% (-2.9%)
480p	93.4%	91.7% (1.7%)	77.3%	66.4% (10.9%)
720p	91.6%	84.5% (7.1%)	73.5%	52.4% (21.1%)
1080p	85.2%	73.7% (11.5%)	65.2%	40.8% (24.4%)
Avg	90.9%	86.2% (4.7%)	72.3%	58.9% (13.4%)

Table 2. The ratio of relative bitrate error on each model and red numbers represent the gain obtained from our proposed method.

Conclusion & Future work

- Main Contributions:
  - A 2nd order CRF-R model that improved 7.8% at the ratio of relative bitrate error within 20% and 27% at bitrate error within 10%.
  - Improved machine learning based on bitstream features reached 90.94% at bitrate error within 20% and 72.25% at bitrate error within 10% averagely.
- In the future, we will utilize decoding buffer pixel domain features to further improve the performance.
- Q & A?

Thank You !