



**HAL**  
open science

# A Case Study of Machine Learning Classifiers for Real-Time Adaptive Resolution Prediction in Video Coding

Madhukar Bhat, Jean-Marc Thiesse, Patrick Le Callet

► **To cite this version:**

Madhukar Bhat, Jean-Marc Thiesse, Patrick Le Callet. A Case Study of Machine Learning Classifiers for Real-Time Adaptive Resolution Prediction in Video Coding. 2020 IEEE International Conference on Multimedia and Expo, Jul 2020, Virtual Conference, United Kingdom. pp.1-6, 10.1109/ICME46284.2020.9102934 . hal-03186419

**HAL Id: hal-03186419**

**<https://hal.science/hal-03186419>**

Submitted on 31 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A CASE STUDY OF MACHINE LEARNING CLASSIFIERS FOR REAL-TIME ADAPTIVE RESOLUTION PREDICTION IN VIDEO CODING

*Madhukar Bhat*<sup>\*†</sup>, *Jean-Marc Thiesse*<sup>\*</sup>, *Patrick Le Callet*<sup>†</sup>

<sup>\*</sup> VITEC, 99 rue Pierre Semard, 92320 CHATILLON, France

<sup>†</sup> LUNAM University, University of Nantes, Group IPI Lab LS2N, Rue Christian Pauc , Nantes 44306

## ABSTRACT

At lower bit-rate encoding video in real-time with a reasonable viewing quality is challenging. Content adaptive per-title encoding is usually leveraged for OTT/VOD delivery by selecting the optimal resolutions and qualities of a given video using multiple encodings. Built on such powerful resolution selection principles, this paper introduces an on the fly resolution prediction without requiring multiple encoding with the help of machine learning which is suitable for real-time video delivery. Two machine learning networks are defined based on the resolution of the previous decision period. Three types of machine learning classifiers: weighted SVM, Random Forests (RF), and custom-designed Multi-Layer Perceptron (MLP) are tested. Suitability of classifiers for real-time resolution prediction is discussed based on the accuracy, BD-rate performances, and impact of misclassification on encoding performance and hardware implementability. The proposed solution offers a promising average bit-rate savings up to 12.6%.

**Index Terms**— Machine learning, Real-time delivery, Resolution prediction, Random Forest(RF), MLP

## 1. INTRODUCTION

Video delivery optimization is thriving in this day and age to address rapid expansion in video consumption. There are several types of video consumption in different scenarios. Applications such as On-demand service, point-to-point communication has been an integral part of modern-day Internet usage. Depending on usage scenarios, real-time (live) video delivery is challenging, considering many network conditions and user consumption methods [1].

Point-to-point video streaming methods deliver video to its members at a given bit-rate in real-time. In this scenario, Constant Bit-Rate (CBR) video encoding methods are applied. Depending on bit-rate, Quantization Parameter (QP) is derived using pre-analysis. Delivery of a video below a specific bit-rate for a particular resolution introduces significant distortion resulting from high QP. One option to achieve acceptable encoding at such resolution is to send a lower number of pixels through resolution or frame-rate down-sampling

[2]. The bit-rate below which switching to a lower resolution is better depends on the temporal and spatial features of video content. The newest video standard in progress Versatile Video Coding (VVC) [3] incorporates an option to select a resolution in a normative way through Reference Picture Resampling (RPR)[4] considering these factors.

Content adaptive encoding techniques have been widely adapted to account for network conditions and video characteristics. Typically, adaptive streaming techniques such as DASH [5] and HLS [6] provide a multi-representation delivery framework for video depending on network conditions. In [7], recommendations are provided on ranges of bit-rates for different resolutions for the OTT scenario. While these applications focus on general thumb rule, it is asserted that some contents can have a different complexity of scenes that might behave differently.

Professional solutions such as [8] are already leveraging such principle, based on the creation of a bit-rate/resolution ladder for various video profiles maximizing quality for VOD applications. Netflix, a popular OTT content provider, proposed dynamic optimizer in video coding [9]: each scene is encoded at multiple down-scaled resolutions and a range of bit-rates. The convex hull of bit-rate/qualities is produced to select the best operating resolution at a given scene using metrics such as PSNR and VMAF [10] after up-scaling all encoded videos to the full resolution. In [11], deep video pre-coding is performed with a design of CNN based multiple downscaling filter network operating before video transmission. Resolution selection is carried out based on the pre-coding network-based single RD point of different scales. It has reduced complexity compared to Netflix based optimizer, which needs several rate-distortion points per scale factor.

The above methods encode the source video at many different resolutions and bit-rates before transmission. Then RD-based selection of the best quality video at a given network condition is selected. These techniques are designed for on-demand delivery. In the case of a point-to-point live delivery system, it would be computationally expensive and introduce unmanageable latency to encode the video at different bit-rates and then select the best resolution based on the convex hull.

There are solutions [12][13] based on scalable video cod-

ing (SVC) [14] by encoding multiple layers of resolution to deliver video at suitable resolution for different network conditions. In these methods, instead of optimizing the encoding resolution, the SVC framework is used for the selection of suitable resolution based on general thumb rule for delivery.

To practically find better-performing resolution for a real-time video delivery system, particularly at a given lower bit-rate scenario, a potential solution would be to have a priority decision based on pre-analysis to predict the best encoding resolution for the decision period. Considering potential VVC's support of switching to a lower resolution, it is more relevant than ever to have a simple but effective method guiding switching to lower resolution with pre-analysis of initial frames in the live point-to-point video coding scenario.

This paper proposes a novel on the fly adaptive resolution prediction for a decision period before encoding process without the RD-score requirement for each resolution allowing the saving of multiple resolution encodings. Then, computationally less expensive real-time and hardware friendly machine learning, weighted SVM, RF's, and custom-designed MLP, are studied. Features from the first few frames are used to predict better performing resolution for a decision period. In this paper, we consider two resolutions, the original resolution, and down-scaled resolution by a factor of 4 in terms of area. This prediction is interpreted as a binary classification problem in the context of machine learning. Depending on the resolution of the previous decision period, two separate classifiers Network-HR and Network-LR, are employed. During training, a weight-based label is proposed to avoid misclassification resulting in quality loss. A professional HEVC encoder is employed for testing in this paper. Labeling can be formulated for any codec, including VVC in the future. Accuracy of classifiers on these algorithms is discussed, and respective BD-rate comparison to reference without adaptive resolution selection is carried out. Finally, the impact of misclassification is analyzed.

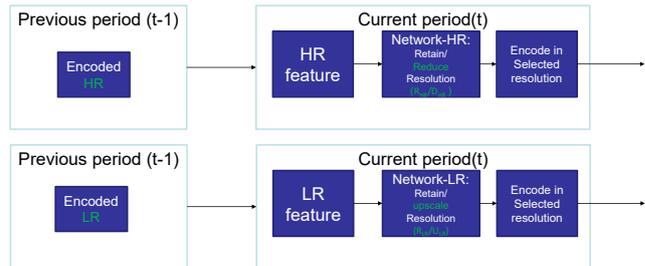
The paper is organized as follows. Section 2 details the Real-time resolution selection algorithm. Section 3 explains the learning algorithms and optimization process. The results are discussed in Section 4. Section 5 details the conclusion drawn.

## 2. REAL-TIME RESOLUTION SELECTION

This section introduces the proposed adaptive real-time resolution selection as a binary classification problem. Two machine learning networks are employed, depending on the resolution of the previous decision period. One of these networks predicts a suitable resolution for the current decision period.

### 2.1. Resolution selection as a classification problem

Prediction of better resolution on the fly for a decision period in a video is challenging. The difficulty arises mainly due



**Fig. 1.** Two machine learning networks are used based on the resolution of previous decision period

to the unavailability of every frame of a decision period of a video for pre-analysis. A general thumb-rule in terms of bit-rate, QP, or any other parameter for resolution selection would miss content-based adaptation. Hence, alternative solutions such as machine learning algorithms are better to approach to predict operating resolution for the current scene in real-time scenarios.

Key questions addressed in this paper are when to change the resolution, how often to check for resolution change, and how to change resolution. Real-time implementations in HEVC, CBR encoding mode, employs a speedy real-time compatible pre-analysis step to guide the encoding process. Features derived from pre-analysis before the main encoder begins for the current decision period is utilized for predicting better resolution. On how often to change resolution, a caveat needs to be addressed for codecs that do not support normative Reference Picture Resampling. Codecs other than VVC consider resolution change as starting of new bit-streams. The decision period is set as the Intra period or scene change in this paper to avoid potential playback issues. The downsampling filter applied in this paper is a professional three tap-linear filter. Since the upscaling happens at decoder end and the majority would do linear filtering, the same is assumed for ground truth formation for the dataset.

### 2.2. Machine learning Networks for classification

In this paper, two machine learning networks are proposed. One of the networks is used depending on the resolution of the previous decision period. Fig. 1 describes these networks. Two machine learning networks are used mainly to address potential adaptive rate-control interference. This interference would potentially happen if we assume original resolution at the starting of each decision period and then decide to switch to a lower resolution or not. Consider lower resolution is selected multiple times in a row, utilization of encoder pre-analysis data used for resolution selection from original resolution has to be re-scaled or re-calculated. Two learning networks are used so that the re-calculation of rate control pre-analysis is required only during resolution change.

### 2.2.1. Resolution selection mechanism

If the previous decision period resolution was original (HR) resolution, learning Network-HR is employed. Network-HR works as a classifier, which could predict to retain resolution ( $R_{HR}$ ) or downscale the resolution ( $D_{HR}$ ). Learning Network-LR is employed if the previous decision period resolution was the downscaled resolution. Network-LR would work as a classifier that would predict to retain lower resolution ( $R_{LR}$ ) or upscale to a higher resolution ( $U_{LR}$ ).

### 2.2.2. Classification labels

The selection of best resolution for the training process, i.e., label formation required for supervised classification methods, could be made in many ways. One way is to measure the quality (Q) of upscaled LR and HR against the source and select a better one for the label ( $L_i$ ) formation in-terms of PSNR. Label  $L_i$ , for a sample, is given as,

$$L_i = \begin{cases} 0, & \text{if } Q_{HR_i} - Q_{LR_i} \leq 0 \\ 1, & \text{if } Q_{HR_i} - Q_{LR_i} > 0 \end{cases} \quad (1)$$

## 3. LEARNING ALGORITHM DESIGN

In this section, dataset creation and training processes for learning algorithms are detailed.

### 3.1. Training and testing samples

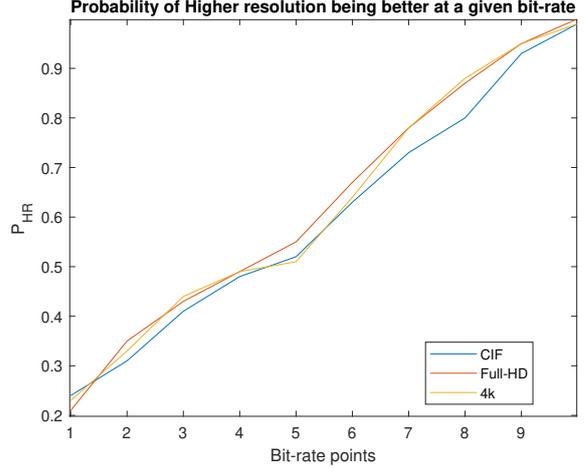
Data selection for training samples for a machine learning algorithm needs to be plentiful and diverse. In the case of resolution prediction, only one sample is available for each Intra period. Compared to other machine learning proposals for video compression, which mostly operate at CTU level, the amount of samples available for resolution selection is sparse by orders of magnitudes. With the consideration of data sparsity, videos are collected from Xiph [15] for CIF, and Full HD and open-source Vimeo videos listed in [16] are selected for both full HD and 4K.

	CIF	Full-HD	4K
Bit-rates	2-20 kbps	0.2-1 Mbps	1-5Mbps

**Table 1.** Bit-rate ranges for CIF, Full-HD and 4K

For a specific operating resolution, bit-rates are selected in such a way that there is a benefit in resolution selection. Fig. 2 shows the probability of retaining original resolution of video  $p_{HR}$  for bit-rates given in Table 1. It shows selected bit-rates are most useful for each operating resolution in-terms of having the ability to choose LR or HR.

Dataset creation, training, and testing are performed separately for each resolution. The selected Intra period is 2 seconds. Labels are formed at each Intra period. A total of 8



**Fig. 2.** Probability of retaining higher resolution at tested bit-rates. Point 2-9 in x-axis are in the range of selected bit rates in Table 1. Point 10 and 1 are 10% higher and lower bit-rates than point 2 and 9 respectively

different bit-rates are encoded for each video in the specified bit-rate range. Training samples are pre-processed to make the dataset at each resolution balanced for each class. In total, 9832 samples for CIF, 4464 samples for Full-HD, and 3624 samples for 4K are trained.

In addition, testing samples are created for measuring the performance of machine learning algorithms. The following sequences are tested for each resolution. For CIF, SVT videos [17] combined to form 2000 frames with 5 different scenes is tested. For full-HD SVT video combined to form 1000 frames with five different scenes, three full-HD videos from JCT-VC common conditions [18] are tested. For 4K videos, five videos from JCT-VC common conditions are tested.

### 3.2. Feature selection

Features selected for a learning algorithm needs to be representative of the purpose. In this paper, it is an adaptive resolution prediction for a decision period. It is important to select minimal features that successfully select a better resolution to avoid an increase in complexity leading to an implementation difficulty in the real-time encoding scenario. The objective is to use features from the first few frames to decide a resolution for the whole decision period. Available features are evaluated as follows.

#### 3.2.1. Evaluation of available features

Evaluation of useful features can be availed through methods such as a wrapper method [19] based on F-score of the network for each and combinations of features, filter models [20] such as mutual information [21] and correlation and regularization of model during training process [22].

In this paper, mutual information is used at first to minimize massive data such as motion vector, the histogram of pixels, and variance since its much faster than the wrapper method to select features. Then F-score for SVM is calculated for reduced features to verify the usefulness of features. After this analysis, the final selected features are listed below. They are divided into four categories Rate control based, Spatial features, Temporal features, Encoder pre-analysis based features. In total, 43 features are selected.

**Rate-Control Based features:** This section of features is relevant to the rate control algorithm to keep the encoded bit-stream under targeted bit-rate.

- $X_{Rate-control-features}$ : Three features Initial QP, Frame rate and targeted bit-rate

**Spatial features:** Spatial features such as histogram of pixel values and variance of CTU's of Intra frame are taken into account with mutual information in consideration.

- $X_{var}andX_{hist}$ : Maximum, minimum and average of Variance of all CTU's and histogram of all pixels of the Intra frame respectively

**Temporal features:** Temporal features are needed to determine the temporal complexity of the scene. Motion vectors which can be made available from pre-analysis are extracted. Besides, the scene change index of each frame of the first GOP in the decision period is taken into account. The scene change index is derived from fast motion estimation of lowered resolution of the source video.

- $X_{Mvx}$  and  $X_{Mvy}$ : Motion vectors in Horizontal and vertical direction of middle picture in first GOP for 16x16 block size. Ten maximum and the average motion vector in both horizontal and vertical directions are selected.
- $X_{scene-change-score}$ : Scene change score of each frame.

**Encoder pre-analysis based features:** For the CBR mode of encoding, HEVC encoders have a very fast pre-analysis where QP estimation for all pictures in a GOP and overall probability of Intra is predicted.

- $X_{QP-est}$ : Estimated QP of all frames in the initial GOP
- $X_{p-intra}$ : Estimated Intra probability

### 3.3. Training process

A training process aims to maximize the accuracy of the classification rate on the test set of a training dataset. The training process for best resolution prediction for the decision period/scene is tested with three techniques: linear SVM, RF, and Multi-layer Perceptron (MLP) classifier. These three learning techniques are computationally less expensive and require no additional hardware requirement. Section 6 shows the performance of these classifiers with an angle to hardware implementation. The impact of misclassification on PSNR is considered as the sample weight to improve accuracy.

#### 3.3.1. Sample weight generation for mis-classification

The impact on PSNR arising from misclassification is assessed using the difference in PSNR. Let R/C be the misclassification when the classifier chooses to retain resolution R instead of changing resolution C compared. In such case weight for that sample  $W_{si}$  is defined as,

$$W_{si} = \frac{R_{PSNR} - C_{PSNR}}{R_{PSNR}} * \frac{C_{rate}}{R_{rate}} \quad (2)$$

This weight account for the varying effect of misclassification and successfully avoids misclassification with very high PSNR differences. The ratio of bit-rate is to account for bit-rate difference induced by encoder control compared to the target rate. In case of Network-HR misclassification is retaining resolution instead of downscaling or vice versa ( $R_{HR}/D_{HR}$ ), ( $D_{HR}/R_{HR}$ ) and for Network-LR it is retaining resolution instead of upscaling and vice versa ( $R_{LR}/U_{LR}$ ), ( $U_{LR}/R_{LR}$ ). In the next subsections design of learning algorithms is given in detail.

#### 3.3.2. SVM and Random Forests

SVM tries to find a unique separating hyper-plane that maximizes the margin between two classes. In this paper, a non-linear SVM is utilized with a polynomial kernel with a degree of 2, and a kernel coefficient of  $1/n\_features$  is applied.

RF classifiers employ many single correlated decision trees, and the final decision is made by gathering results from all single decision trees and de-correlating them. In this paper, RF classifiers are used since they are lightweight and have a better trade-off with accuracy and over-fitting training data. The configuration adopted in the context of resolution prediction is: tree depth of 12, and the number of estimators is 25.

#### 3.3.3. Multi-layer Perceptron (MLP) classifier design

The proposed MLP is designed in such a way that it is hardware friendly. For hardware, it is desirable to have similar mathematical operations for decisions coming from different networks. Hence, the target is to use the identical network topology for all resolution and both Network HR and LR. With this in mind and considering design guidelines are given in [23], an MLP network with four fully connected hidden layers with neuron sizes (15,11,7,5) was selected for both Network-HR and Network-LR. The activation function used was the Rectified Linear Unit (RELU). The applied loss function was the mean squared error with sample weights derived in the last section. L1 regularization for the first layer was added to avoid over-fitting. Adam (Adaptive Moment Estimation) optimizer was adopted for weight update as it requires less memory and its easiness in handling hyper-parameter optimization. For hardware implementations MLP are better than RF as the same architecture for hardware computation

remains for all resolutions. On the other hand, an RF classifier might have a different architecture for each update of a dataset and each resolution.

## 4. RESULTS

In this section, the performance of learning algorithms explained in the previous section is discussed in two contexts. At first, the accuracy of networks is analyzed. Then the BD-rate performance of each network and the impact of misclassification on the performance of encoding is studied.

### 4.1. Accuracy of Networks

One can measure the accuracy of machine learning networks by calculating precision (P), Recall (R), and F1-score (F1) [24]. Results for SVM, Random-forest, and MLP networks for each resolution are reported in Table 2. Each selected network achieved excellent overall accuracy. RF classifier outperforms both SVM and MLP in terms of its accuracy mainly due to similar topology constraint put-forward for MLP across all conditions.

Conditions	Precision(%)			Recall(%)			F1-score(%)		
	CIF	FHD	4K	CIF	FHD	4K	CIF	FHD	4K
<b>SVM</b>									
Network-HR	79	75	75	81	75	75	80	75	75
Network-LR	79	76	72	80	76	75	79	76	74
<b>RF</b>									
Network-HR	<b>89</b>	<b>89</b>	<b>87</b>	82	<b>89</b>	<b>87</b>	85	<b>89</b>	<b>87</b>
Network-LR	<b>85</b>	<b>86</b>	<b>87</b>	<b>82</b>	86	85	<b>83</b>	<b>86</b>	86
<b>MLP</b>									
Network-HR	86	85	<b>87</b>	<b>84</b>	<b>87</b>	<b>87</b>	<b>86</b>	86	<b>87</b>
Network-LR	80	82	<b>87</b>	80	<b>89</b>	<b>87</b>	80	85	<b>87</b>

**Table 2.** Accuracy of the network for both learning network-HR and LR based on resolution of previous period

### 4.2. BD-Rate performance

In this section, the BD-rate performance of each machine learning algorithm is tested against the reference encoding of high resolution. A comparison to ground truth resolution is reported in Table 3. Test sequences mentioned in the Section 3 are tested for BD-rate performance. Bit-rates tested were four bit-rates in the range of 5-20kbps for CIF, 0.1-1Mbps for Full-HD, and 1-5Mbps for 4K videos. Table 3 shows that dynamic resolution selection introduces bit-rate gain at the same quality. The potential to gain bit-rate at the same quality increases with the resolution. Overall BD-rate performance of MLP algorithm and RF are close, but SVM has lower BD-rate savings because of its lower accuracy. In some sequences, the BD-rate savings of learning algorithms are higher than ground truth. This is because, in some cases, even-though wrong decisions are made bit-rate of LR is much smaller than HR, and the difference in quality is small.

Bit-rate saving /Sequence	Ground truth	MLP	SVM	RF
<b>CIF</b>				
SVT	-3.61%	-3.11%	-3.12%	-3.15%
Park joy	-6.62%	-5.57%	-4.79%	-5.9%
Crowdrun1	-0.62%	-0.61%	-0.6%	-0.61%
Crowdrun2	-1.03%	-1.02%	-0.89%	-1.02%
Flags	-1.01%	-1.01%	-1.01%	-1.01%
<i>Overall</i>	-2.58%	-2.27%	-2.08%	-2.33%
<b>Full-HD</b>				
SVT	-2.52%	-2.46%	-1.02%	-2.46%
ParkJoy	-3.8%	-3.8%	-0.42%	-3.8%
Basket_Ball	-3.2%	-1.38%	-1.38%	-1.38%
Crowdrun	-3.74%	-4.08%	-2.71%	-3.87%
<i>Overall</i>	-3.31%	-2.93%	-1.38%	-2.88%
<b>4K-sequences</b>				
Catrobot	-3.25%	-3.25%	-1.54%	-3.06%
Tango	-13.26%	-13.26%	-13.26%	-13.26%
Campfire	-8.73%	-6.69%	-6.01%	-7.01%
Foodmarket	-14.16%	-14.16%	-9.5%	-14.16%
Rollercoster	-25.61%	-25.61%	-8.09%	-21.28%
<i>Overall</i>	-13%	-12.59%	-7.52%	-11.75%

**Table 3.** BD-rate at same quality compared to single higher resolution at CIF, Full-HD and 4K

### 4.3. Impact of misclassifications

Misclassification of the trained network in resolution switch incurs a loss of quality compared to the ground truth. In the context of this paper, the main quality loss that is to be studied is when the lower resolution was falsely selected instead of the original resolution. This sub-section aims at analyzing maximum loss against a reference with only high resolution. The condition for this occurrence would be the false positive in Network-HR and false negative in Network-LR. In this analysis, maximum PSNR loss ( $dPSNR(C/R)_{max}$ ) and BD-rate loss ( $BD_{loss}(C/R)_{max}$ ) due to misclassification in tested sequences are reported in Table 4.

Misclassification losses	MLP	SVM	RF
<b>CIF</b>			
$dPSNR(C/R)_{max}$ (dB)	0.24	0.24	0.17
$BD_{loss}(C/R)_{max}$	8.52%	8.52%	8.01%
<b>Full-HD</b>			
$dPSNR(C/R)_{max}$ (dB)	0.6	0.63	0.59
$BD_{loss}(C/R)_{max}$	8.45%	9.91%	8.14%
<b>4K-sequences</b>			
$dPSNR(C/R)_{max}$ (dB)	0.26	1.12	0.26
$BD_{loss}(C/R)_{max}$	3.11%	24.51%	3.11%

**Table 4.** Maximum PSNR loss and Bit-rate loss during misclassification

Overall performance analysis shows that MLP and RF classifiers perform better than SVM in resolution prediction at the starting of an intra-period. RF classifier sometimes edges MLP mainly due to similar architecture constraints put forward in designing MLP for its hardware friendliness. The same topology would have the same template of a code port for efficient weight multiplication for prediction. As mentioned earlier, RF classifier could pose challenges in hardware

implementation by changing its tree architecture in case of updating algorithm with more data, which would require a change of the code suitable for hardware implementation.

## 5. CONCLUSION

In the context of real-time content-adaptive video encoding, this paper introduces an on the fly resolution prediction suitable for lower bit-rate scenarios where multiple encoding for selecting better resolution is not desirable. In the proposed framework, machine learning classifiers SVM, RF, and designed MLP are studied for different video resolutions. Depending on the resolution of the previous Intra-period, two machine learning networks are defined. Original resolution and downsampled resolution by a factor of 4 in terms of the area are tested in this paper. The performance of Learning algorithms has been evaluated in terms of accuracy, BD-rate, and misclassification impact for CIF, Full-HD, and 4K resolution. MLP and RF classifiers are better than SVM with MLP more suitable for hardware implementations. Adaptive resolution prediction using MLP introduces an average bit-rate gain of 2.3% - 12.6% depending on the resolution of the video.

## References

- [1] M. K. Mukerjee, D. Naylor, J. Jiang, D. Han, S. Seshan, and H. Zhang, "Practical, real-time centralized control for cdn-based live video delivery," *ACM SIGCOMM Computer Communication Review*, vol. 45, no. 4, pp. 311–324, 2015.
- [2] A. Segall, M. Elad, P. Milanfar, R. Webb, and C. Fogg, "Improved high-definition video by encoding at an intermediate resolution," in *Visual Communications and Image Processing 2004*, International Society for Optics and Photonics, vol. 5308, 2004, pp. 1007–1018.
- [3] J.-R. Ohm and G. J. Sullivan, "Versatile video coding—towards the next generation of video compression," in *Picture Coding Symposium 2018*, 2018.
- [4] J. Luo, V. Seregin, and W. Wan, "Summary report on reference picture resampling filters," in *JVET-P0021-v3*, 2019.
- [5] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," *IEEE multimedia*, vol. 18, no. 4, pp. 62–67, 2011.
- [6] W. M. R.P. Pantos, *Http live streaming draft-pantos-http-live-streaming-13*, Network Working Group: 1., 2016.
- [7] D. Weinberger, *Choosing the right video bitrate for streaming hls and dash*, 2018.
- [8] ATEME, *The single critical factor for high-quality streaming, now and forever*, White paper, 2019.
- [9] I. Katsavounidis, *Dynamic optimizer — a perceptual video encoding optimization framework*, The Netflix Tech Blog, 2018.
- [10] Z. Li, A. Norkin, and A. Aaron, "Vmaf-video quality metric alternative to psnr," *Joint Video Exploration Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11*, 2016.
- [11] E. Bourtsoulatze, A. Chadha, I. Fadeev, V. Giotsas, and Y. Andreopoulos, "Deep video precoding," *arXiv preprint arXiv:1908.00812*, 2019.
- [12] M. Wien, R. Cazoulat, A. Graffunder, A. Hutter, and P. Amon, "Real-time system for adaptive video streaming based on svc," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 9, pp. 1227–1237, 2007.
- [13] O. Abboud, K. Pussep, A. Kovacevic, and R. Steinmetz, "Quality adaptive peer-to-peer streaming using scalable video coding," in *IFIP/IEEE International Conference on Management of Multimedia Networks and Services*, Springer, 2009, pp. 41–54.
- [14] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the h. 264/avc standard," *To appear in IEEE Transactions on Circuits and Systems for Video Technology*, p. 1, 2007.
- [15] Xiph, *Xiph.org video test media*, [derf's collection].
- [16] D. Vatolin and D. Kulikov, *Hvc video codecs comparison 2018*, Enterprise edition, 2018.
- [17] L. Haglund, "The svt high definition multi format test set," *Swedish Television Stockholm*, 2006.
- [18] JCTVC-L1100, "Common test conditions and software reference configurations," 2013.
- [19] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.
- [20] H. Liu and H. Motoda, *Computational methods of feature selection*. CRC Press, 2007.
- [21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 8, pp. 1226–1238, 2005.
- [22] S. Ma and J. Huang, "Penalized feature selection and classification in bioinformatics," *Briefings in bioinformatics*, vol. 9, no. 5, pp. 392–403, 2008.
- [23] S. Karsoliya, "Approximating number of hidden layer neurons in multiple hidden layer bpnn architecture," *International Journal of Engineering Trends and Technology*, vol. 3, no. 6, pp. 714–717, 2012.
- [24] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *European Conference on Information Retrieval*, Springer, 2005, pp. 345–359.