Network State Estimation by Analyzing Raw Video Data

Yimeng Sun[a] Badr Mochizuki[a]

st191919@m1.kcg.edu b_mochizuki@kcg.edu

[a]School of Applied Information Technology, The Kyoto College of Graduate Studies for Informatics, Kyoto, Japan

Abstract — With the rapid development of the Internet, video streaming services are playing a bigger role in our daily life. Due to the COVID-19 pandemic in 2020, telework using web camera has become a new working style for most people. In this situation, current networks are subjected to a higher traffic load due to increased bandwidth-consuming video applications. As a result, heavy network congestion may occurs and affect video streaming services negatively. In this study, we focus on estimating the network state by analyzing video data. Raw video data have been collected and input into deep neural network models in order to estimate the bandwidth and package loss ratio of the network. Numerical results show that our model achieved a high MAE value for bandwidth prediction of 4.44. As for packet loss ratio, we have obtained a MAE value of 0.04 using PSNR. Finally, we believe that MAE can be increased further in our future work.

Keywords— Neural network, video classification, network state estimation, video quality, PSNR.

I. EXPERIMENTS

A. Class handling

Original videos are transmitted to an imaginary video viewer via a network emulator. Dataset for training is generated based on network conditions as follows:

- Bandwidth range: from 1100 kbps to 2000 kbps at 100 kbps intervals.
- Packet loss ratios: 0.001%, 0.01%, 0.025%, 0.05%, 0.1%, and 0.25%.

Therefore, there is a total of $10 \times 6 = 60$ classes in all the 480 received videos. We have decided to study bandwidth and packet loss ratio separately because the data size for each class is too small for a machine learning algorithm to be able to learn properly.

B. Bandwidth

In this experiment, original videos are streamed from a video streamer to a video viewer using a Network Emulator (NE) over RTP [1]. This NE has the capability to control traffic such that different bandwidths can be achieved. Consequently, this will affect the throuput of the

Table 1. File size and bit rate gaps between received and original videos.

File name	File size	Bit rate (video)
0GHpTnbnTZs.mp4 (original)	48.3 MB	1702 kbps
0GHpTnbnTZs_1100kbps_0001.mp4	36.4 MB	1380 kbps
0GHpTnbnTZs_1200kbps_0001.mp4	39.9 MB	1513 kbps
0GHpTnbnTZs_1300kbps_0001.mp4	41.3 MB	1566 kbps
0GHpTnbnTZs_1400kbps_0001.mp4	42.6 MB	1616 kbps
0GHpTnbnTZs_1500kbps_0001.mp4	44.5 MB	1689 kbps

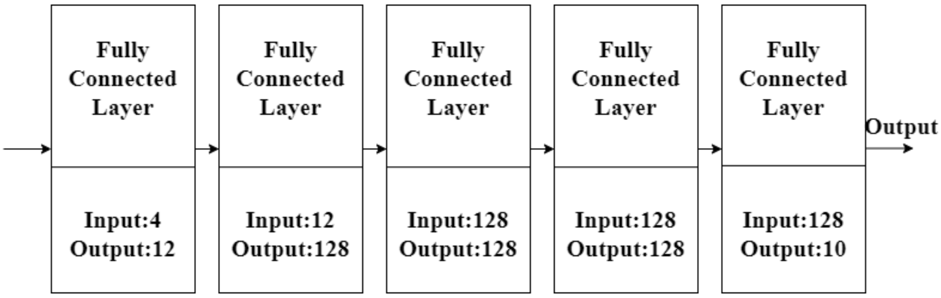

Figure 1. Deep neural network model for predicting bandwidth.

network which will in turn affect the quality of the received videos in terms of loss, bit rate, file size etc. Since there is a correlation of file size and video bit rate between received and original videos, we think that studying these parameters may help us predicting the bandwidth of the received videos. Table 1 shows an example of the file size and bit rate gaps between received and original videos.

However, we have noticed that some videos do not follow this regular pattern (e.g. video evBD9ieSL7g), therefore, we have decided not to use these as training data.

The file size and bit rate of each video have been collected for calculating the difference between original and received ones. These data are input into a deep neural network model along with the original file size and bit rate. The architecture of the deep neural network model is shown in Fig. 1.

C. Packet loss ratio

Transferring videos using different packet loss ratios has also an effect on the metadata of the file but that effect is limited compared to bandwidth. On the other hand, packet loss has a larger and negative effect on the quality of a video because transmitting a video via a network with a high packet loss ratio can cause block noise or glitch in the video, hence deteriorating the quality of the video.

The quality of a video can be evaluated by peak signal-to-noise ratio (PSNR) for each frame of the video [2]. PSNR is often used as a quality measure between an original and a compressed image [3]. Compared to other similarity evaluation methods, PSNR is more robust to minor pixel displacement that often happens in video transfer [4]. PSNR is defined in Eq. (1) as follows:

$$PSNR = 20 \times log_{10}(MAX_I) - 10 \times log_{10}(MSE) \quad (1)$$

Where,

MAX_I is the maximum possible pixel value of the image.

$MAX_I = 2^N - 1$, when the pixels are represented using N bits per sample.

MSE is the mean squared error between the original and the received image.

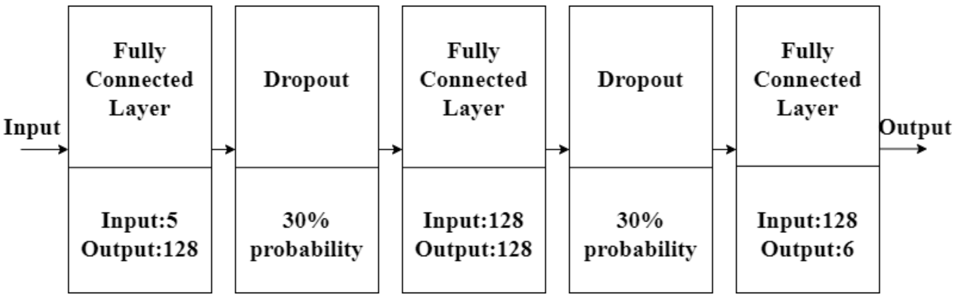

Figure 3. Proposed neural network model for predicting packet loss ratio.

The unit of PSNR is decibel (dB). The higher the PSNR, the better the quality of the received image.

The effect of loss on the quality of a video may be larger or smaller depending on the position of the lost pixel (lost packet) in a particular frame [2]. Therefore, rather than analyzing some specific frames, we have concluded that an overall analysis of the video would be a better choice.

First, every received video is compared to the original one by calculating the PSNR for every frame. After analyzing this data, we have noticed that there is a correlation between the packet loss ratio, the frame count and the PSNR value of each frame.

Original frame Received frame A Received frame B

PSNR between original frame A and received frame A = 19.58

PSNR between original frame B and received frame B = 10.00

Figure 2. PSNRs between frames received in different condition.

According to the correlation matrix, frames with a PSNR value less than or equal to 42, are highly correlated to the loss ratio. Therefore, we have considered only the frames with PSNR values less than or equal to 42. Then, the number of frames with PSNR values less than 42, 41, 40 and 39 have been counted and the percentage for each category was calculated. Finally, the above calculated frame count and percentage has been input into a neural network model. The bandwidth of the video has also been passed as an input to the network. The architecture of the proposed neural network model is shown in Fig. 3.

II. RESULTS AND CONSIDERATION

A. Model evaluation

The performance of the proposed models is evaluated based on the mean absolute error (*MAE*), which is defined in Eq. (2) as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Estimation[i] - Answer| \qquad (2)$$

Where n is the number of tests.

MAE has been calculated for both the bandwidth and the packet loss ratio.

Deep learning prediction models for both bandwidth and packet loss ratio have been evaluated by *MAE*. The results of both models are shown in Table 2 as follows.

Table 2. MAE of models for studying bandwidth and loss ratio.

	Model 1 (bandwidth)	Model 2 (loss ratio)
MAE	4.44	0.04

B. Results

Data from test videos have been collected and input into neural networks using the same method as the training data. Prediction results of test videos are shown as follow

Table 3. Prediction results of test videos.

File name	Bandwidth (predict)	Loss ratio (predict)
0.mp4	1200	0.001
1.mp4	1300	0.001
2.mp4	1300	0.25
3.mp4	1300	0.001
4.mp4	1300	0.25
5.mp4	1100	0.25
6.mp4	1200	0.25
7.mp4	1500	0.25
8.mp4	1600	0.25
9.mp4	1300	0.25

C. Consideration

In this study, we focus on estimating the network state by analyzing video data. We studied the correlation between video metadata and bandwidth, frame PSNR and packet loss ratio. Raw video data have been collected and input into deep neural network models to estimate the bandwidth and package loss ratio. The best mean absolute error (MAE) of bandwidth has a result of 4.44. This means the bandwidth of a video that between 1100 and 2000kbps can be predicted with only an error of 4.44, and we consider it a good method to predict bandwidth from received videos. However, MAE of packet loss ratio has a result of 0.04, meaning the average error in predicting packet loss ratio between 0.01 and 0.25 will be 0.04, which is much higher compared to the method for predicting bandwidth. We believe this method can be improved further in our future work.

III. REFERENCES

[1] E. A. H. Schulzrinne, “RTP: A Transport Protocol for Real-Time Applications,Request for Comments 3550,”July 2003. Available: https://tools.ietf.org/html/rfc3550..

[2] Da-lu Zhang et al., “Analysis the effect of packet loss on video quality of experience,” *Computer Engineering and Applications,* 46(1), pp. 71-71, 2010.

[3] H. A. A. Y. Mochizuki, “Low-Latency Full HD Video Streaming using Wavelet-based Compression on GPU,” Information Processing Society of Japan, 2009.

[4] “a la carte of image quality evaluation,” Available: https://dftalk.jp/?p=18111.