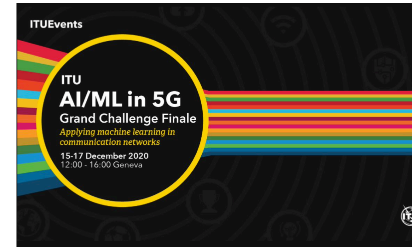


Challenge for estimation of bandwidth and loss rate by focusing on the degradation characteristics of raw video data

[Challenge Title] PS-031- NEC, Japan
Network State Estimation by Analyzing Raw Video Data

Affiliation : Osaka Prefecture University

Team Member : Yuusuke Hashimoto, Yuya Seki, Daish Kondo

Team Name : JOJO

Outline

- Problem Statement
  - Background
  - The goal of this challenge
  - Description of a given dataset
- Proposed Solution
- Performance Evaluation
- Conclusion

Outline

- Problem Statement
  - Background
  - The goal of this challenge
  - Description of a given dataset
- Proposed Solution
- Performance Evaluation
- Conclusion

The background of challenge

- The demand for **interactive live video streaming services** is dramatically growing
  (i.e., Remote work system using web cameras)

The Internet needs to accommodate the dramatic increase in traffic generated by such video streaming services

Conventional Solution

- Previous related works in the field of (non-interactive) video streaming
  → Estimating network state by using playback buffer state and adaptively controlling the bit rate [1][2][3]

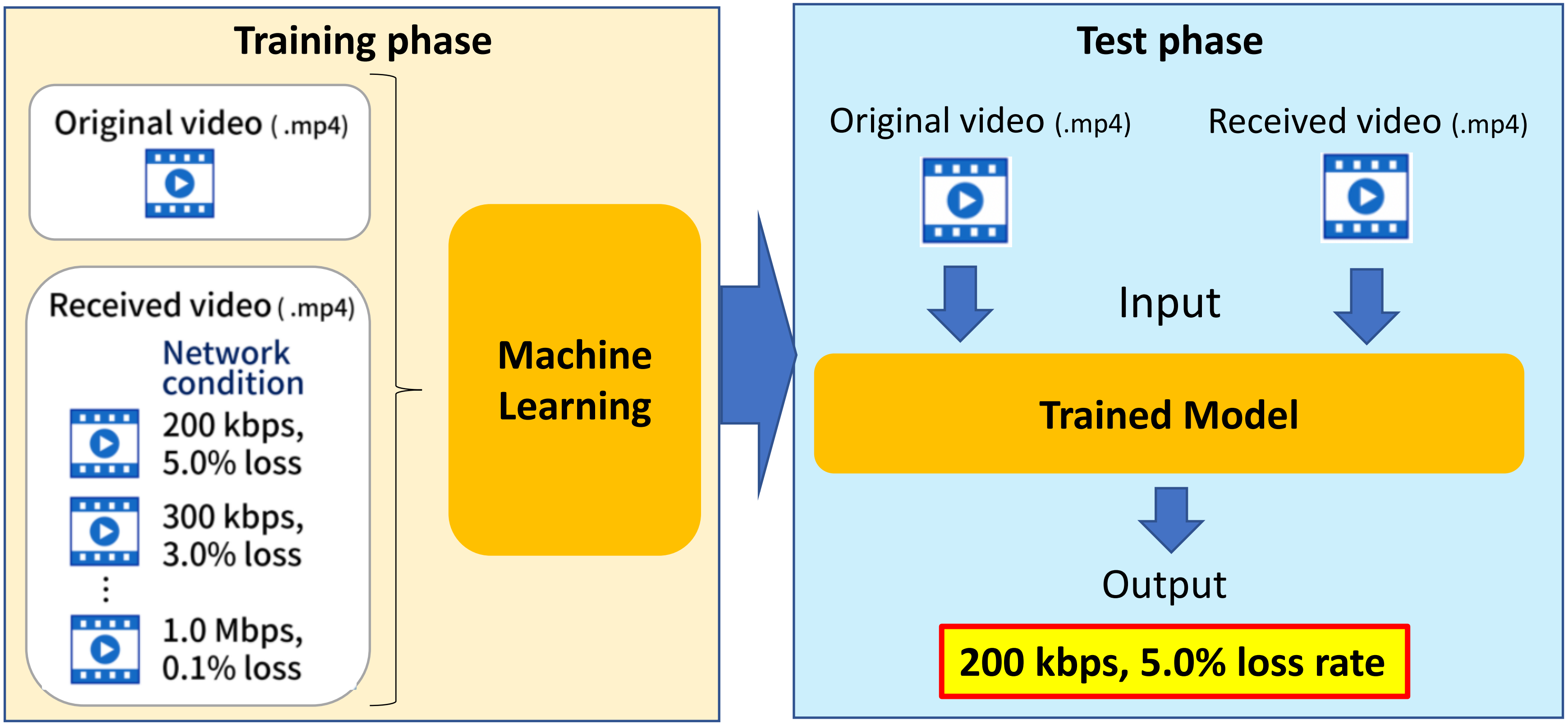
- Strict constraints on interactive live video streaming : real time communication
  - Unable to prefetch video content like non-interactive video streaming
  - Impossible to use playback buffer state on the receiver side

Causing a challenging issue of passive network state estimation by analyzing raw video data

[1]Akamai. 2016. dash.js. https://github.com/Dash-Industry-Forum/dash.js/. (2016).
[2]Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. 2014. A buffer-based approach to rate adaptation: evidence from a large video streaming service. SIGCOMM Comput. Commun. Rev. 44, 4 (October 2014), 187–198.
[3]K. Spiteri, R. Urgaonkar, and R. K. Sitaraman. 2016. BOLA: Near-Optimal Bitrate Adaptation for Online Videos. CoRR abs/1601.06748 (2016).

The goal of this challenge

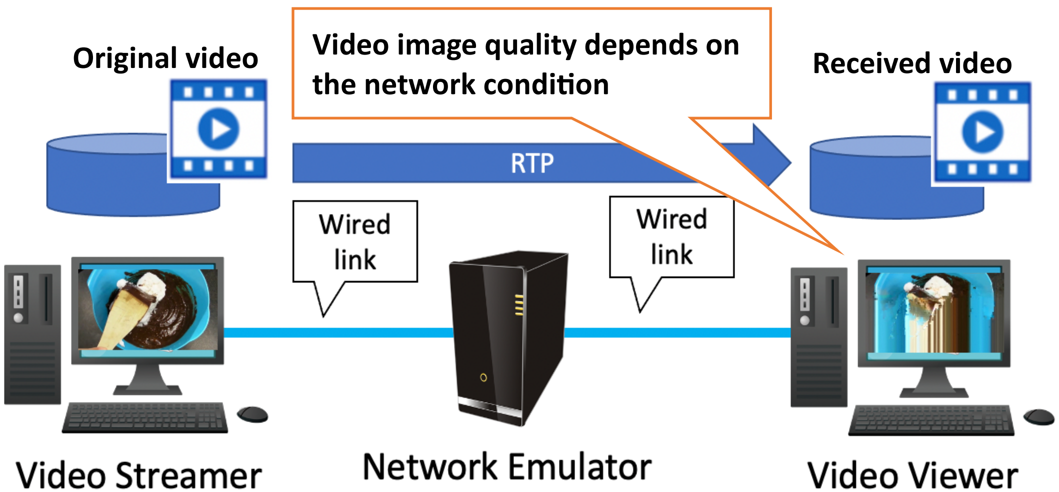
- Estimating network state, i.e., **bandwidth** and **packet loss ratio**, with given raw video data
- Training and testing the AI model using video data labeled with the network state



Description of a given dataset

- Two types of raw video data (.mp4 format)
  - **Original video**
    - open data (YouTube - 8M[4])
  - **Received video**
    - generated in a lab environment as shown in the following figure

> The network emulator controls video traffic with a predefined bandwidth and loses packets at a predefined loss rate.



Pattern of network conditions

Bandwidth

1100kbps , 1200kbps,…,2000kbps (100kbps intervals)

✖

Packet Loss Rate

0.001%, 0.01%, 0.025%, 0.05%, 0.1%, 0.25%

Lab network environment for provided dataset

[4] YouTube - 8M, https://research.google.com/youtube8m/

Outline

- Problem Statement
  - Background
  - The goal of this challenge
  - Description of a given dataset
- Proposed Solution
- Performance Evaluation
- Conclusion

Overview of our solution

Objective

- Estimating network state (bandwidth and packet loss rate) with given raw video data

Proposed solution

- Using time-series data of signal-to-noise ratio (PSNR) for estimating network state
- Determining the main factors of video degradation (bandwidth shortage, packet loss) based on data rate distribution of video data
- Partially extracting PSNR time series data for each determined factor of degradation and using for training the model

Unique Points!

Contribution

- Providing new knowledge to the challenge of clarifying the relationship between raw video data and network state

Our solution

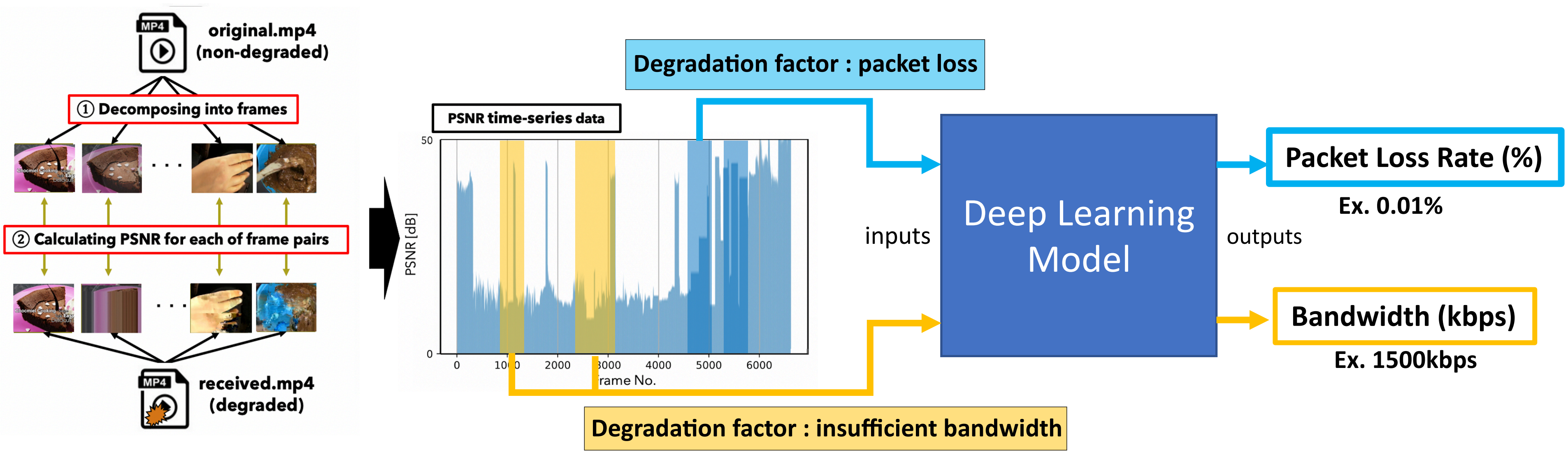
Overview of processing steps

(Step1) Decomposing original video data (.mp4) and received video data (.mp4) into frame units

(Step2) Calculating peak signal-to-noise ratio (PSNR) time-series data by comparing frame pairs

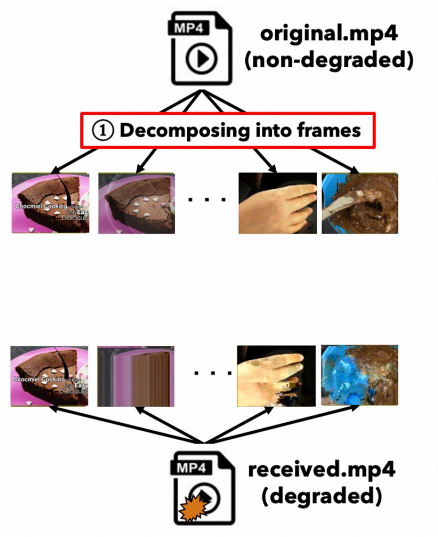
(Step3) Partial estimation of video degradation factors (insufficient bandwidth, packet loss) based on bit rate time series data of video

(Step4) Training AI models using partial PSNR time series data extracted for each degradation factor, and estimating the network state using the trained model



Our solution

(Step1) Decomposing original video data (.mp4) and received video data (.mp4) into frame units



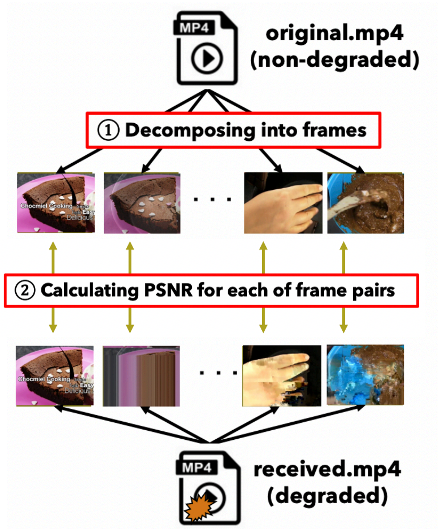
Our solution

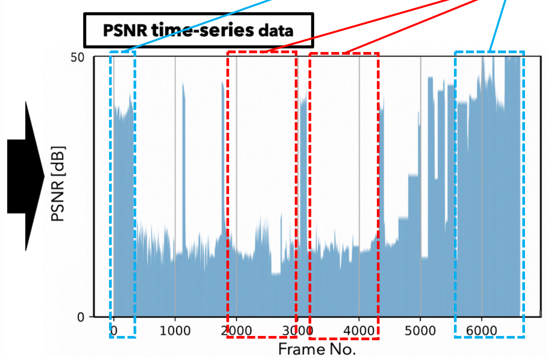
(Step2) Calculating **peak signal-to-noise ratio (PSNR)** time-series data by comparing frame pairs

- The most popular **video quality assessment (VQA)** metrics
- **Full reference (FR)** based VQA metrics :
  Both the original and degraded video are available, and we compare them to each other to estimate how similar the two videos are frame-by-frame

High PSNR area:
→ **Little degradation** in received video data (= high quality)

Low PSNR area:
→ **Heavy degradation** in received video data (= low quality)

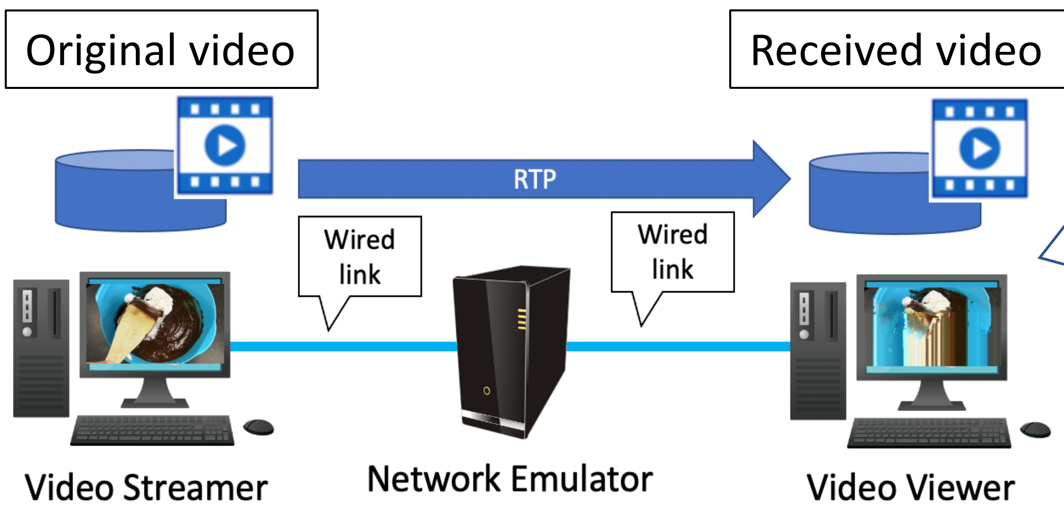


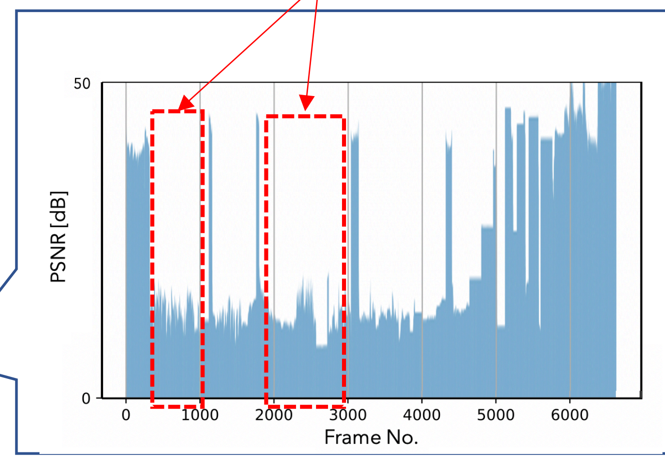


Difficulties in this challenge

- The degradation observed in received video is caused by two degradation factors

Degradation factor:
Insufficient bandwidth? or Packet loss?

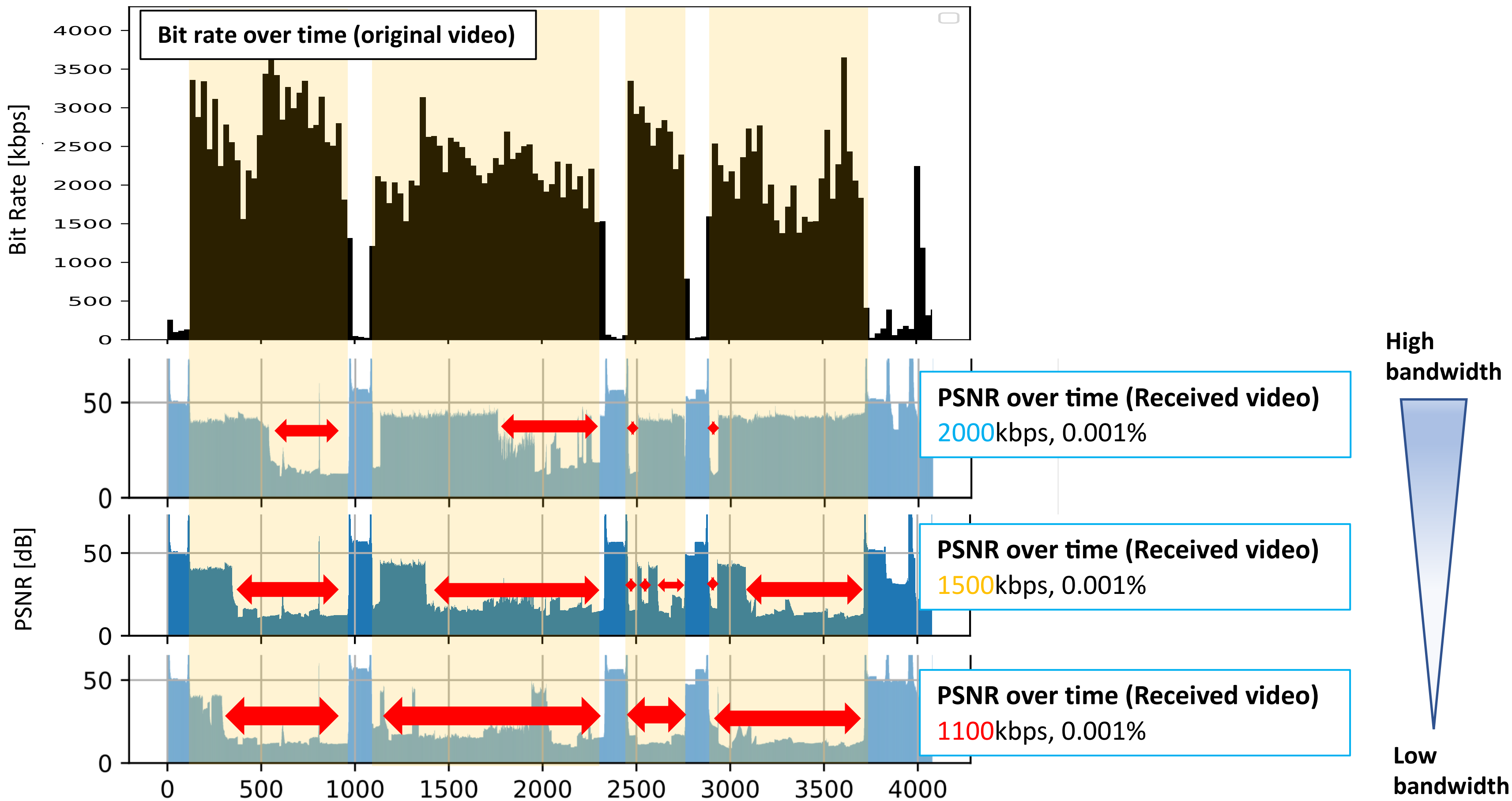




- It is necessary to estimate the factors of degradation in the received video and extract the features of each degradation separately to train the model

The relationship between the bit rate of the original video and the received video image quality

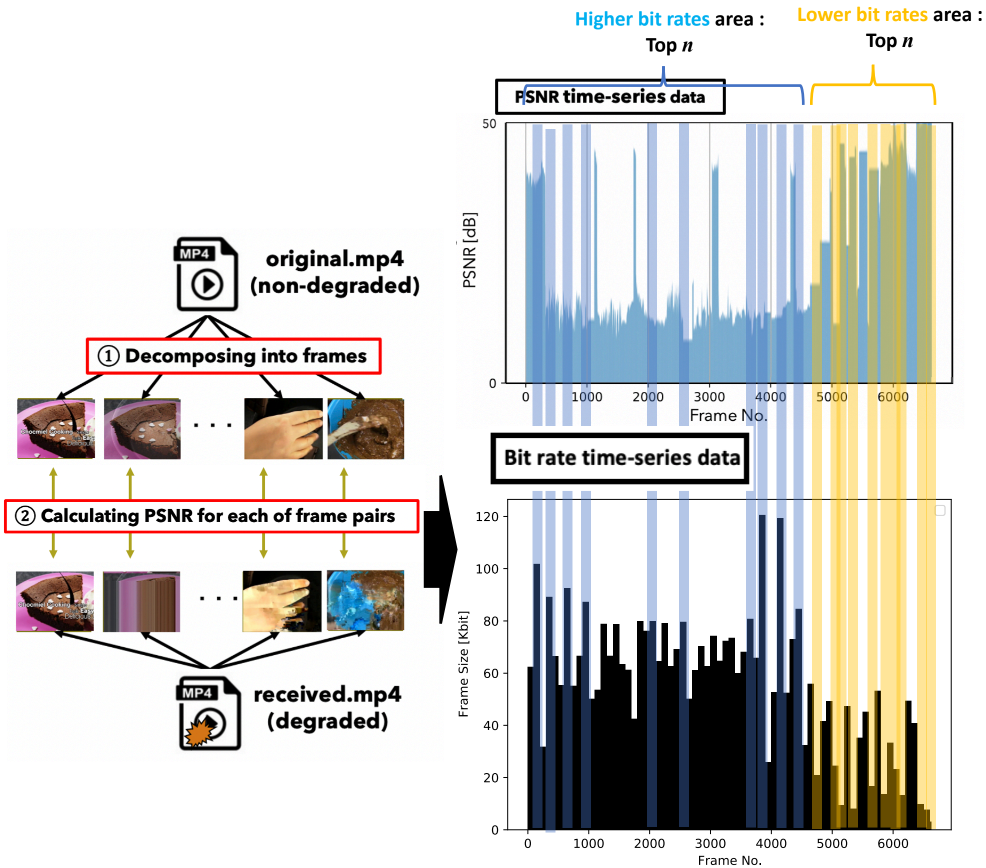
- Comparing the **bit rate over time** with **three PSNR graphs of different bandwidths**



- **[feature 1] Degradation occurs mainly at high data rates in the original video**
- **[feature 2] The smaller the bandwidth is the larger the area of degradation is**

Our solution

(Step3) Partial estimation of video degradation factors (insufficient bandwidth, packet loss) based on bit rate time series data of video

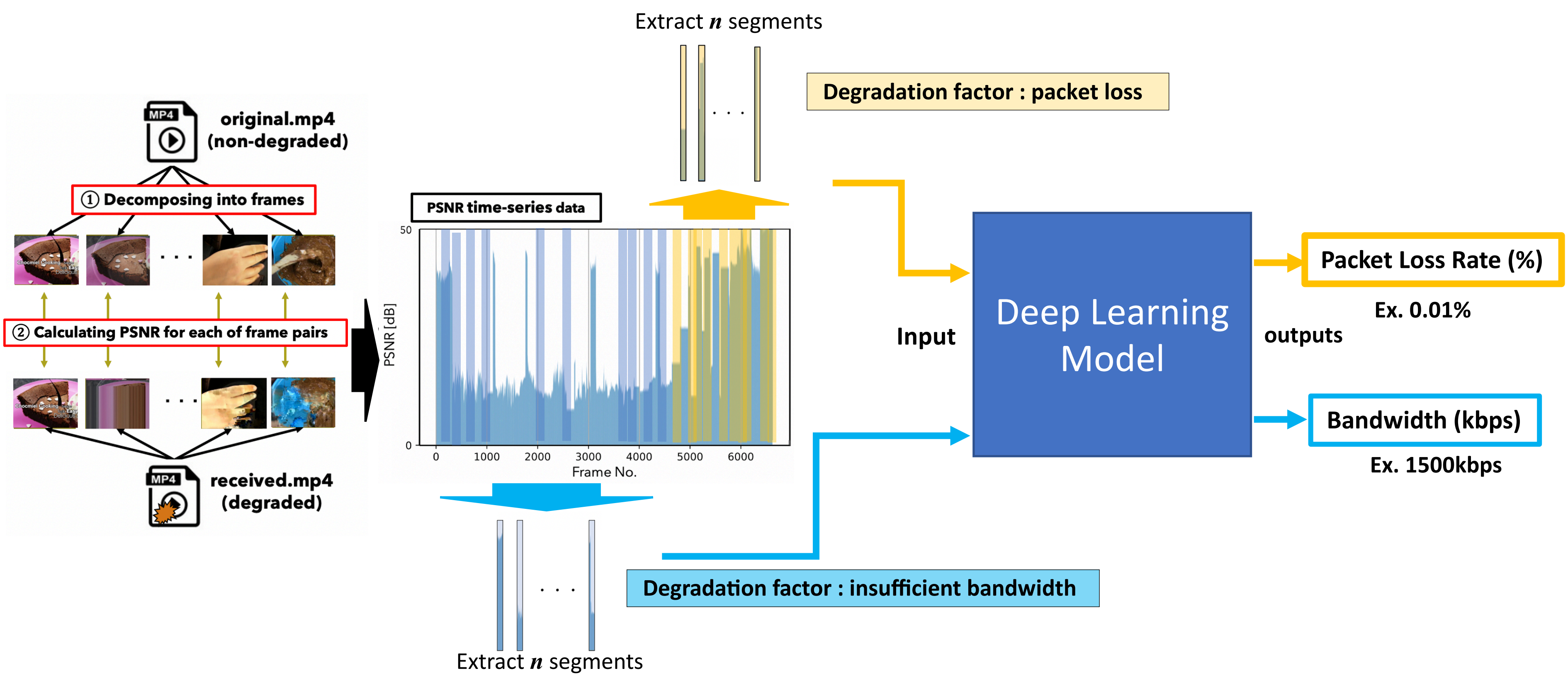


Unique Points

- **Higher bit rates area** : Strongly affected by **bandwidth shortage**
  - ▶ Used for bandwidth estimation
- **Lower bit rates area** : Little or no affect due to **bandwidth shortage** (= **Impact of packet loss is dominant**)
  - ▶ Used for loss rate estimation

Our solution

(Step4) Training AI models using partial PSNR time series data extracted for each degradation factor, and estimating the network state using the trained model

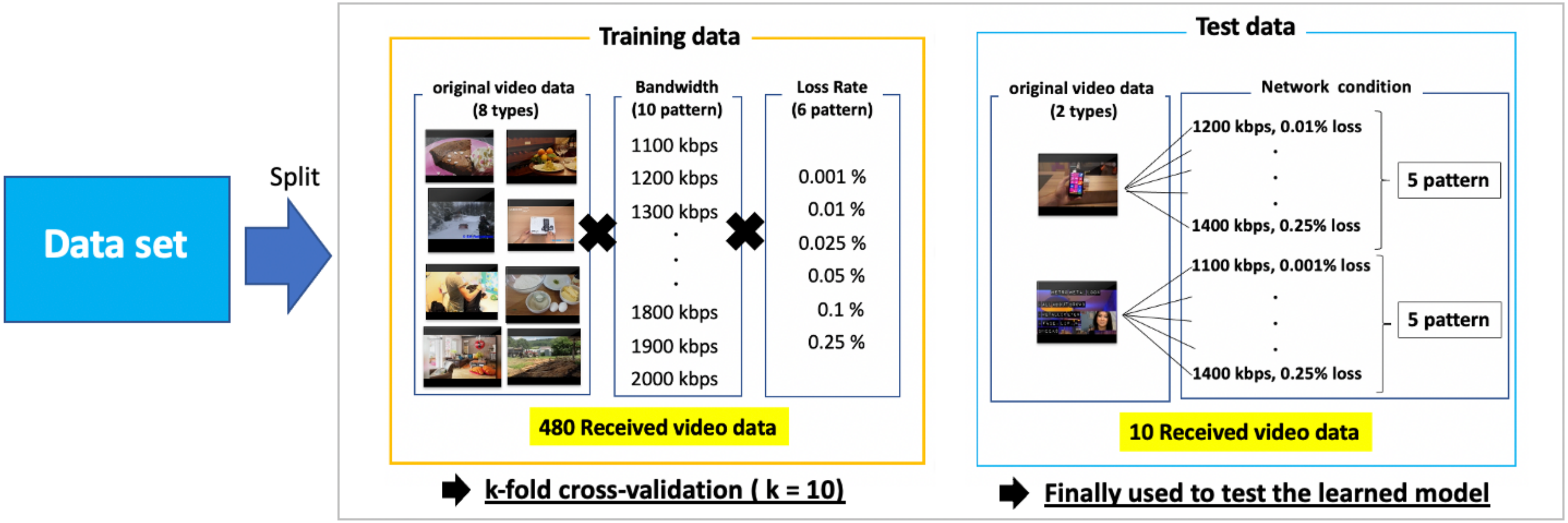


Outline

- Problem Statement
  - Background
  - The goal of this challenge
  - Description of a given dataset
- Proposed Solution
- Performance Evaluation
- Conclusion

Performance Evaluation

- Splitting given raw video data set into training data and test data
- Training and evaluating model using **k-fold cross-validation** (**k = 10**)
- Finally testing the learned model using the test data



- Mini Batch Learning (Batch Size: 16), Epoch: 30
- The number of segments to be extracted n: 50
- Number of PSNR data contained in each segment : For 30 frames
- Loss function : MSE (Mean Squared Error)
- Evaluation function : MAE (Mean Absolute Error)

- Evaluating the accuracy of model by calculating the **MAE (Mean Absolute Error)** of the estimated value and the ground true label

Result of Evaluation

- Results of evaluating the accuracy of the model using k-fold cross-validation (k = 10)

model (1～k)	MAE for Bandwidth [kbps]	MAE for Loss rate [%]
1	208.782	0.06024
2	242.761	0.07309
3	248.685	0.06790
4	237.717	0.06767
5	238.670	0.07594
6	198.426	0.06779
7	205.294	0.05395
8	228.298	0.05883
9	242.593	0.07060
10	289.381	0.06194

Average of 10 MAEs

- Bandwidth : 234.061 , Loss rate :0.06579

Testing the model using test data

- Results of bandwidth and loss rate estimation using 10 test data

	Bandwidth [kbps]		Loss rate [%]	
Test data ID	Predicted value	Ground true label	Predicted value	Ground true label
1	1221.06	1200	0.07798	0.01
2	968.370	1800	0.06882	0.001
3	1545.81	1400	0.07108	0.25
4	1790.07	1600	0.04684	0.025
5	1545.81	1400	0.07108	0.25
6	1213.76	1100	0.07875	0.001
7	1600.88	1300	0.08134	0.001
8	1826.53	1700	0.07702	0.025
9	1857.11	1900	0.06883	0.025
10	1552.01	1400	0.08159	0.25

- MAE of bandwidth : 207.049 MAE of loss rate :0.09378

- Result of bandwidth estimation

Achieving relatively good estimation accuracy (Average Error : Less than 18.8%)

- Result of loss rate estimation

Estimation accuracy is poor considering the order of the ground true labels in the loss rate (e.g. 10^{-3} %)

Discussion

- Why is the accuracy of loss rate estimates significantly worse?

Our Hypothesis

- The effect of bandwidth shortage is less pronounced during times of low video data rates, and only the effect of packet loss is more pronounced.

incorrect

Packet loss features may not be obtained from the extracted samples

- Packet Loss: Probabilistic (not certain)
- Time periods when the data rate is very low in video
  → The same picture in succession with <u>little or no movement</u>

Resilient to quality degradation because inter-frame prediction allows restoration of degraded portions from adjacent frames.

We need to consider further approaches to extract the features of packet loss

Conclusion

Challenge

- Clarifying the relationship between raw video data and network state
- Estimating network state (bandwidth and packet loss rate) with raw video data using machine learning

Proposed Solution

- Analyzing the raw video data using machine learning and estimating the two types of network state
- Determining the main factors of video degradation (bandwidth shortage, packet loss) based on data rate distribution of video data
- Partially extracting PSNR time series data for each determined factor of degradation and using for training the model

Performance Evaluation

- Bandwidth estimation: achieving relatively high estimation accuracy
- Loss ratio estimation: low estimation accuracy (room for improvement)

Future Challenges

- Improving the accuracy of loss rate estimation
- Consideration of the framework with real-world use cases (e.g. latency, feasibility)