ITU AI/ML in 5G Challenge Grand Challenge Finale (Final Conference)

Theme 2 from NEC:

A Lightweight deep learning model for network state estimation using raw video data

Team KCGI
Yimeng Sun , Badr Mochizuki

CONTENTS

Background

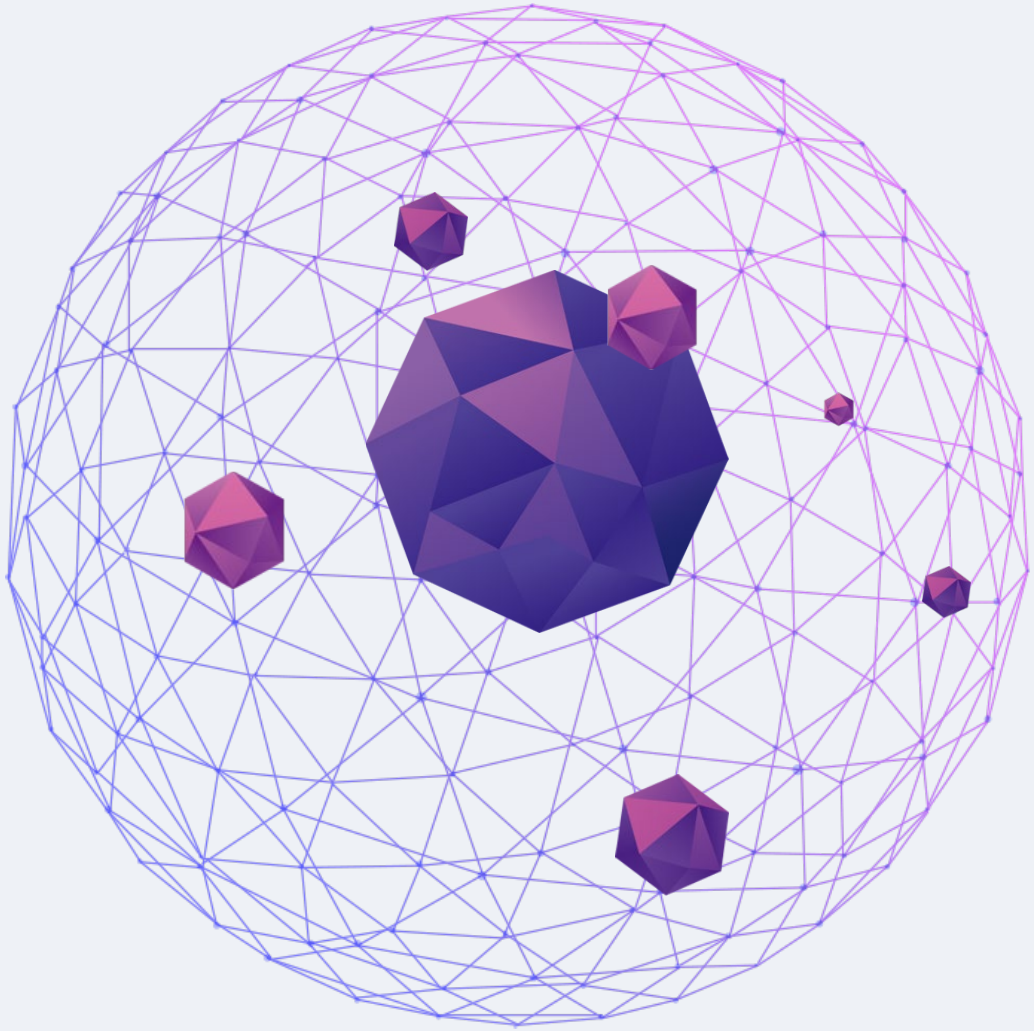
Problem statement

Proposed Method

Conclusion

1

Background



Video streaming services

Webex Meetings

Needs of interactive live video streaming services

Webex Meetings

Google Meet

Survey: Office worker in Tokyo who works from home by telework.*

December 2019

Approximately 15.6%

April 2020

Approximately 49.1%

Interactive live video streaming services (e.g., telework system using web cameras) requires real-time playback that can't buffer videos like YouTube or Netflix.

video quality should be optimized based on the network conditions.

*: Survey by Japan Telework Association

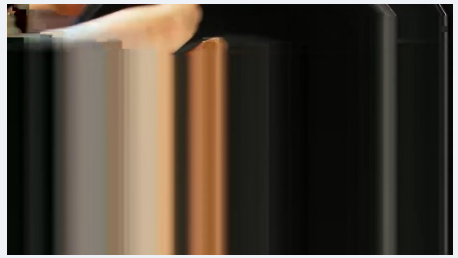
Relationship between network state & video quality

Sender → Network → Receiver

Both over-the-top (OTT) providers and users are sensitive to network conditions because it affects the quality of videos heavily.

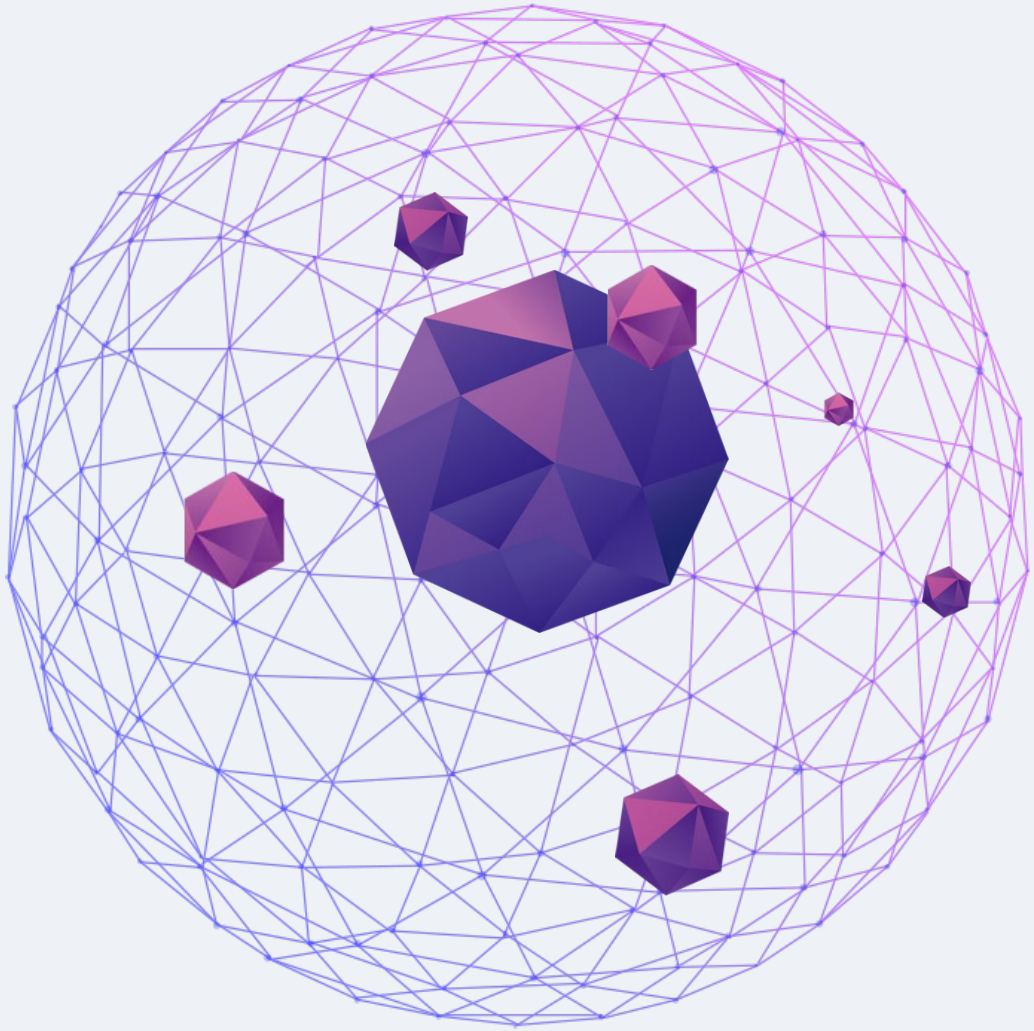
Good condition → **High quality**

Bad condition → **Low quality**



2

Problem statement



Understanding network state from raw videos

Sender → Network → Receiver

Open video dataset

Live video streaming via RTP

with adjustable:

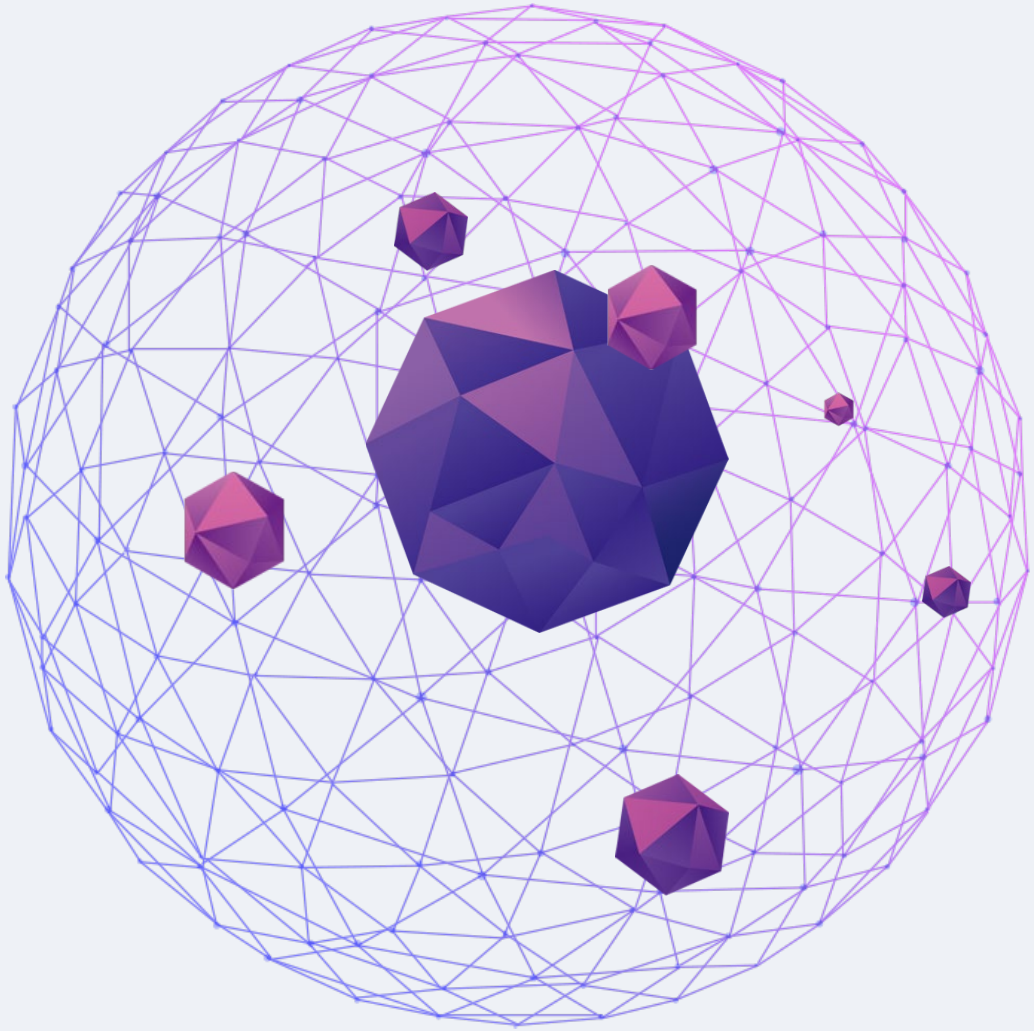
- **Throughput**
- **Packet loss ratio**

Received video data

Task: Estimate throughput/packet loss ratio using video data via ML based method.

3

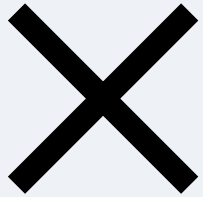
Proposed Method



Classes handling

Throughput(kbps)

- **1100**
- **1200**
- **1300**
- **1400**
- **1500**
- **1600**
- **1700**
- **1800**
- **1900**
- **2000**



Loss(%)

- **0.001**
- **0.01**
- **0.025**
- **0.05**
- **0.1**
- **0.25**

60 Classes

The size of dataset (480) is too small for 60 classes.

Learn throughput and packet loss separately.

Learning approach

Total Classes

- Throughput: 10 classes
- Packet loss ratio: 6 classes

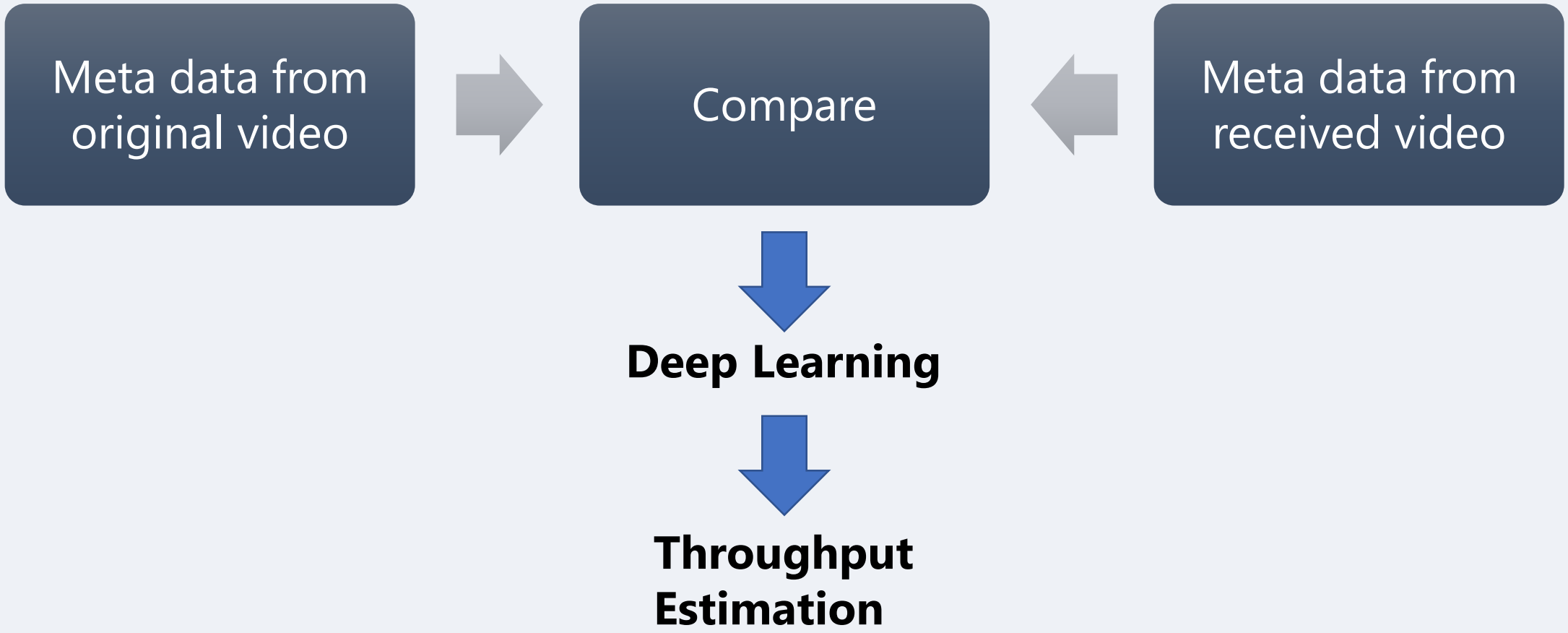
Learning

- Learning throughput without using the data of packet loss ratio.
- Learning packet loss ratio using the data of throughput.

Estimation

- Estimate throughput.
- Estimate packet loss ratio using derived throughput.

Throughput prediction



Comparing File size & Bitrate

Throughput may have a correlation with

- File size of received video
- Video bit rate of received video

Supervised learning : Comparing File size & Bitrate

Learning Data

x: input

name	size	bitrate	size_diff	bitrate_diff
0GHpTnbnTZs_1100kbps_0001.mp4	36.4	1381	11.9	321
0GHpTnbnTZs_1100kbps_001.mp4	37.3	1414	11.0	288
0GHpTnbnTZs_1100kbps_0025.mp4	37.5	1419	10.8	283
0GHpTnbnTZs_1100kbps_005.mp4	38.0	1440	10.3	262
0GHpTnbnTZs_1100kbps_01.mp4	37.2	1411	11.1	291

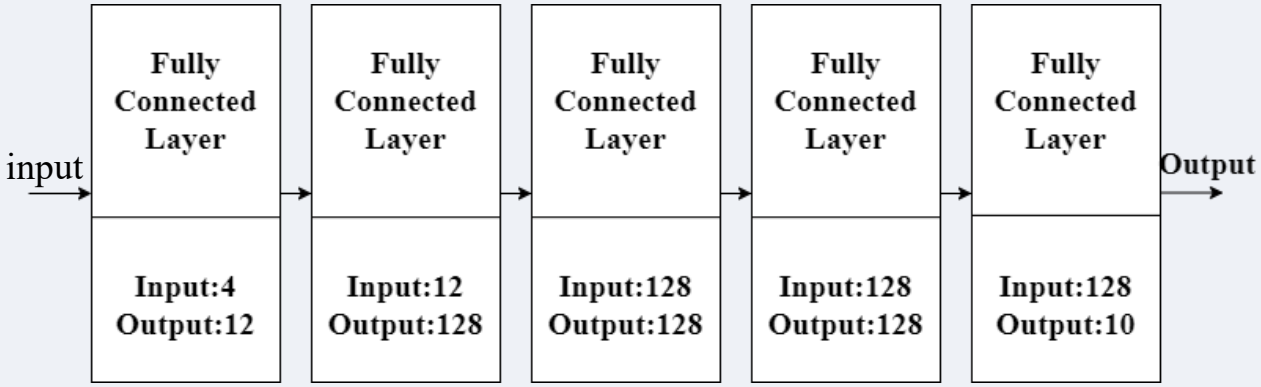
y: label

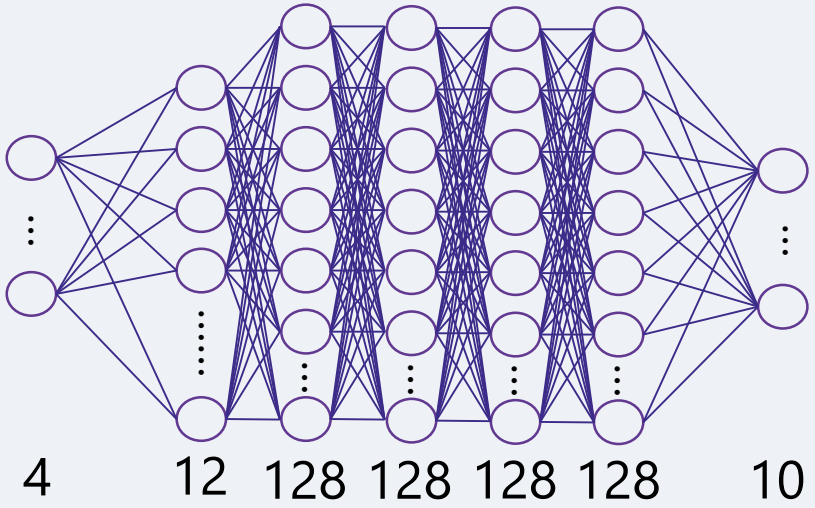
Cls
0
0
0
0
0

- Size: Size of received video(MB).
- Bitrate: Bitrate of received video(Kbps).
- Size_diff: Difference between original video and received video in size.
- Bitrate_diff: Difference between original video and received video in Bitrate.
- Cls: Classes (number between 0-10,which represents throughput between 1100-2000)

Neural Network and Result

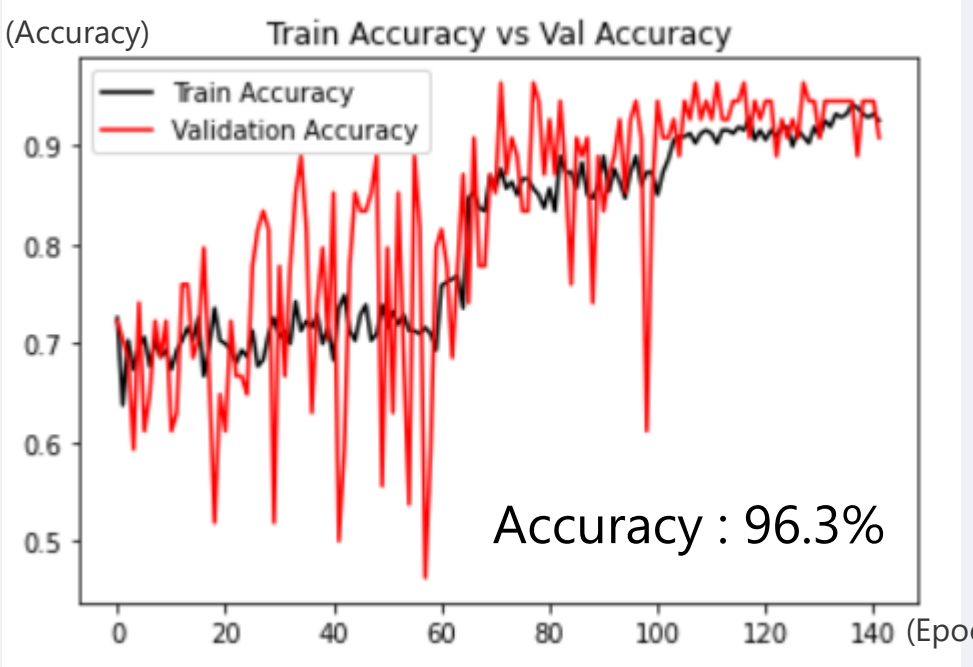
Network Model





Training method

1. Training roughly
   - Epochs = 250, Batch size = 3
2. Training carefully
   - Epochs = 500, Batch size = 3
   - Reduce learning rate on plateau
   - Early stopping

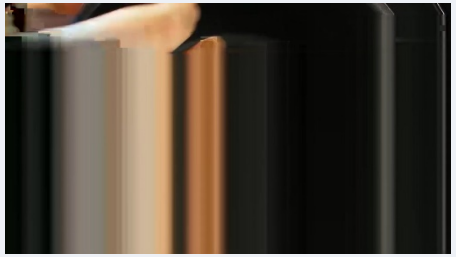


Packet loss ration prediction

Frame extraction: Original video

Compare

Frame extraction: Received video



Throughput

Deep Learning

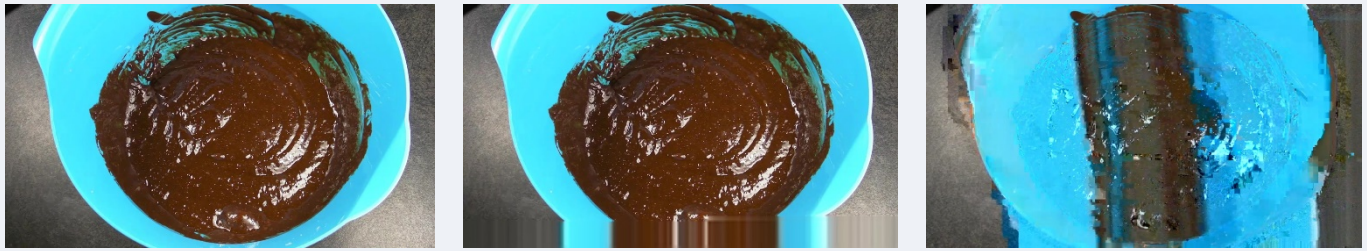
Packet loss Ratio Estimation

Frame Comparison : PSNR[2]

PSNR - a quality measure that is robust for small pixel displacement, which often happens in video transfer.

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

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.



Original A B

PSNR between original frame and received frame A = 19.58
PSNR between original frame and received frame B = 10.00

1. PSNR value of each frame between original one and received one is calculated for all frames in each video.
2. Total number of frames whose PSNR values are less than some specific numbers are being counted and used as input for our model.

Supervised learning : Packet Loss Ratio

Learning Data

	x: input				y: label
name	Condition	42	41	40	39	lossclass
0GHpTnbnTZs_1100kbps_0001.mp4	1100	6074	6045	5818	5709	0
0GHpTnbnTZs_1100kbps_001.mp4	1100	6048	5861	5793	5681	1
0GHpTnbnTZs_1100kbps_0025.mp4	1100	6159	6128	5912	5815	2
0GHpTnbnTZs_1100kbps_005.mp4	1100	5883	5837	5708	5626	3
0GHpTnbnTZs_1100kbps_01.mp4	1100	6092	5903	5855	5744	4

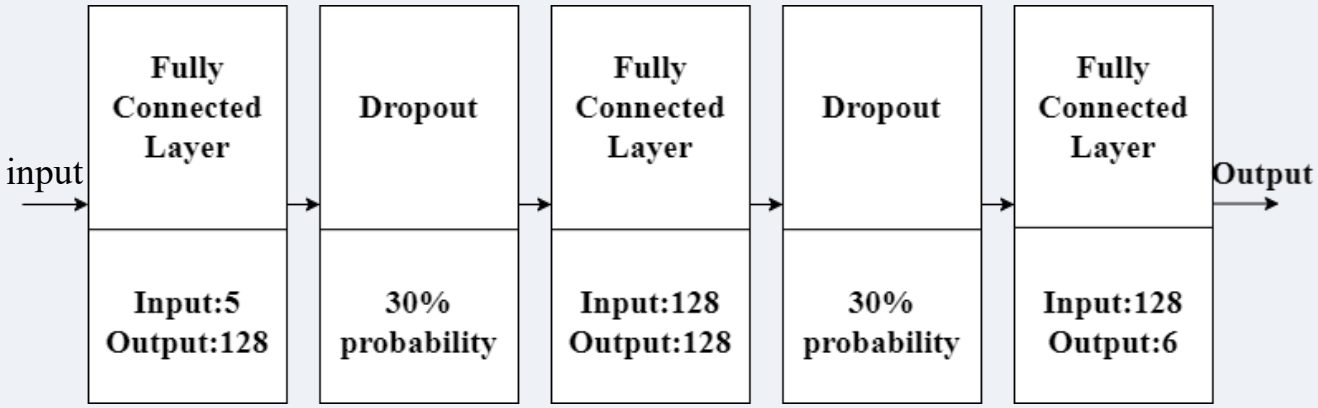
- Condition: Throughput
- 42: Frame count which its PSNR less than 42.
- 41: Frame count which its PSNR less than 41.
- 40: Frame count which its PSNR less than 40.
- 39: Frame count which its PSNR less than 39.

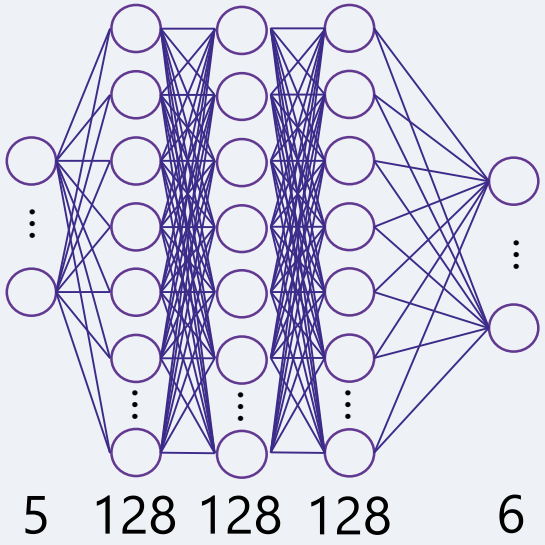
Have a higher correlation than other PSNR thresholds.

- Lossclass: Classes (number between 0-5,which represents packet loss ratio between 0.001%-0.25%)

Neural Network and Result

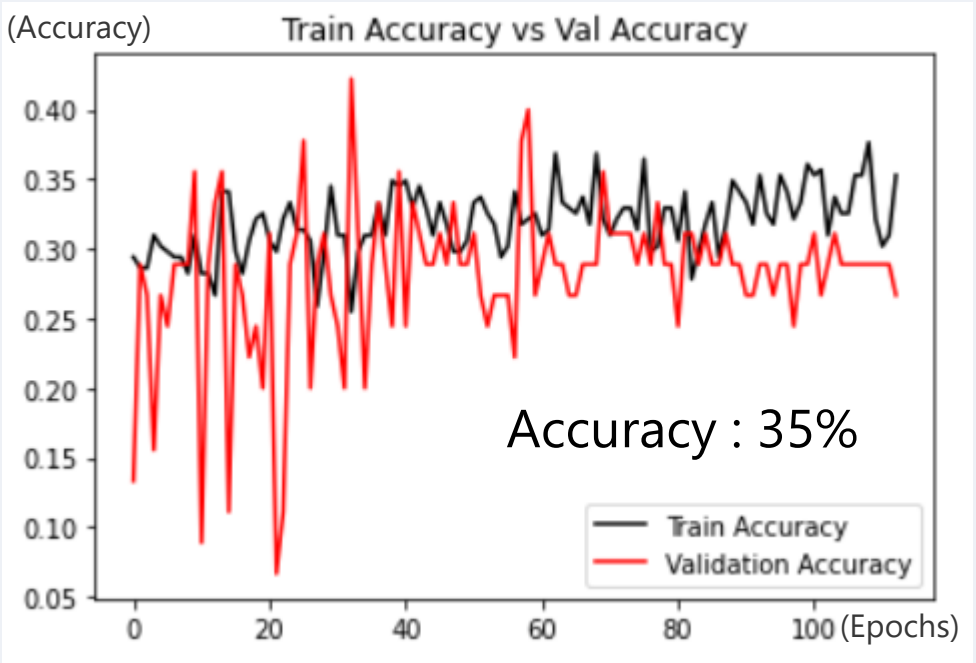
Network Model





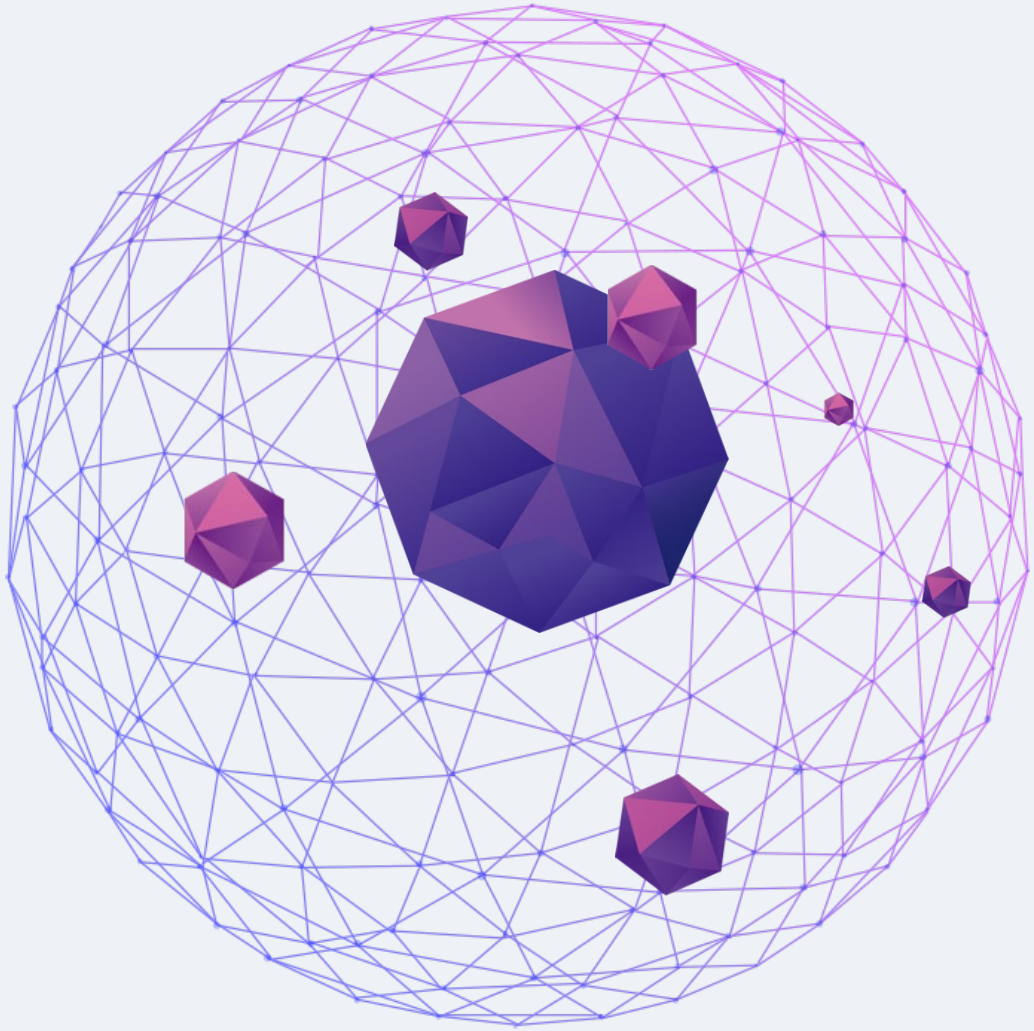
Training method

1. Training roughly
   - Epochs = 250, Batch size = 3
2. Training carefully
   - Epochs = 500, Batch size = 2
   - Reduce learning rate on plateau
   - Early stopping



4

Conclusion



Evaluation

	Model 1 (Throughput)	Model 2 (loss ratio)
Prediction Range	1100~2000(kbps)	0.001~0.25(%)
MAE	4.44(kbps)	0.04(%)

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Estimation[i] - True\ Value[i]|$$

- The best mean absolute error (MAE) of Throughput has a result of 4.44. We consider it a good method to predict throughput from received videos.
- MAE of packet loss ratio has a result of 0.04.
  We believe this method can be improved further in our future work.

Selling Point: advantages of the proposed method

- A precise method to predict throughput from received videos.
- Very lightweight model that can run on almost every platform.
- Only need file size and video bitrate for Estimating throughput.
- Extremely fast: Can estimate throughput for 500 videos in 1 second.

References

[1]	E. A. H. Schulzrinne, “RTP: A Transport Protocol for Real-Time Applications, Request for Comments 3550”, July 2003. https://tools.ietf.org/html/rfc3550.
[2]	D. 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. Ando, Y. Mochizuki, “Low-Latency Full HD Video Streaming using Wavelet-based Compression on GPU”, Information Processing Society of Japan, 2009(3), pp. 1-6, 2009.
[4]	“A la carte of image quality evaluation,” Available: https://dftalk.jp/?p=18111.

Thank you for listening!