

# 스케일러블 비디오 전송을 위한 딥러닝 기반 비트율 및 화질 예측 모델

하 호 진\*

## Deep Learning-based Bitrate and Video Quality Prediction Model for Scalable Video Transmission

Hojin Ha\*

---

본 논문은 2023년도 1학기 한양여자대학교 교내연구비에 의하여 연구됨

---

### 요 약

COVID-19 팬데믹 이후로, 비대면 화상회의와 원격 교육의 증가로 인해 비디오 전송 시스템에 대한 관심이 증가하고 있다. 본 논문은 스케일러블 비디오 부호화기의 입력 파라미터를 활용하여 부호화될 비트스트림의 비트율과 화질을 예측하는 인공 신경망(ANN, Artificial neural network) 모델을 제안한다. 제안된 ANN 모델은 부호화기의 압축 비트율을 실제 부호화과정 없이, 입력 파라미터만을 이용하여 예측할 수 있기 때문에, 네트워크 대역폭과 전송 비트율의 불일치로 발생하는 전송 지연 및 손실을 최소화할 수 있다. 제안된 신경망 모델의 성능평가를 위해서, R2 메트릭을 이용하였으며, 1에 가까운 값을 나타내어 정확도가 높은 예측 모델임을 확인하였다.

### Abstract

Since the COVID-19 pandemic, there has been an increasing interest in video transmission systems due to the rise of remote video conferencing and online education. This paper proposes an Artificial Neural Network(ANN) model that utilizes the input parameters of a scalable video encoder to predict the bit rate and quality of the encoded bitstream. The proposed model enables the prediction of compression bit rates of the encoder without the actual encoding process, thus minimizing the transmission delays and losses caused by network bandwidth and transmission rate discrepancies. The performance evaluation of the proposed neural network model utilizes the R2 metric and we confirmed its high accuracy since the predicting values are close to one.

### Keywords

video transmission system, artificial neural network, ANN, prediction model, scalable video coding, SVC

---

\* 한양여자대학교 소프트웨어융합과 교수  
- ORCID: <http://orcid.org/0000-0001-6488-9193>

· Received: Aug. 01, 2023, Revised: Dec. 01, 2023, Accepted: Dec. 04, 2023  
· Corresponding Author: Hojin Ha  
Dept. Software Convergence, Hanyang Women's University, 200 Salgoji-gil, Sengdong-gu, Seoul, 04763, Korea  
Tel.: +82-2-2290-2205, Email: [hjha@hywoman.ac.kr](mailto:hjha@hywoman.ac.kr)

## 1. 서 론

COVID-19 팬데믹 이후로, 비대면의 화상회의 및 원격 교육이 증가하면서 비디오 전송 시스템에 대한 관심이 증가되고 있다[1]. 일반적으로 비디오 데이터는 높은 네트워크 대역폭을 요구한다. 특히 동시에 여러명의 사용자가 비디오 전송 시스템에 접속했을 때, 더 많은 대역폭을 요구하게 된다. 본 논문에서는 비디오 전송 시스템의 안정적인 서비스를 위해서 스케일러블 비디오 코딩(SVC, Scalable Video Coding) 부호화 방법을 이용한다[2][3]. SVC 부호화 방법은 비디오를 여러 계층(Layer)으로 분할하여, 개별적으로 부호화하기 때문에, 추가적인 송신단의 부하없이 수신단의 성능과 네트워크 대역폭에 맞추어 적절한 비디오 품질을 제공할 수 있는 장점이 있다.

최근에는 스케일러블 비디오의 안정적인 서비스를 위해서 다양한 비디오 전송 방법들이 제안되었다[4]-[9]. 5G 무선 네트워크에서 SVC와 캐싱 기술을 결합하여 비디오 스트리밍 품질을 향상시키는 방법이 제안되었다[4]. 네트워크의 효율성을 높이고 대역폭 요구 사항을 줄여 사용자 경험을 개선하였다. [5]에서는 스케일러블 비디오의 기저계층과 향상계층의 안정적인 전송을 위해, 다중 경로 라우팅 방법을 제안하였다. 데이터량이 많은 비디오 데이터의 전송시에, 발생하는 지연을 최소화하기 위한 중요도에 따라서 다중 경로를 설정하는 최적화 알고리즘을 제안하였다. 최근에는 360 영상의 비디오 스트리밍을 위해서 딥러닝 기반의 적응적 비트율 조절 알고리즘이 제안되었다. 360 영상의 뷰포트 전환시에 급격히 증가하는 비트율을 최소화하기 위해, 딥러닝기반의 강화학습을 통해서, 네트워크 대역폭에 최적의 전송 비트율 조정하는 알고리즘을 제안하였다[6]-[9].

그러나 제안된 스트리밍 방법은 급변하는 네트워크 환경에서, 송신단의 비디오 부호화기가 적절한 전송량을 갖는 비트스트림 생성하여 전송해야 한다. 그러나 전송량을 알기 위해서는 비디오 시퀀스를 부호화해야 실제 전송량을 알 수 있기 때문에, 이로 인한 지연이 발생된다. 본 논문에서는 비디오 부호

화기의 설정 파라미터를 이용하여 실제 비디오를 부호화를 하지 않고, 비트량을 예측할 수 있는 딥러닝(DL, Deep Learning) 기반의 인공 신경망(ANN, Artificial Neural Network) 예측 모델을 제안한다. 최근에 많은 관심을 받고 있는 DL기술은 대용량 데이터의 이용, 높은 계산 능력, 그리고 강력한 인공 신경망 알고리즘을 활용하여, 컴퓨터 비전 및 음성 인식, 자연어 처리, 모델링 등 다양한 분야에서 많이 이용되고 있다[10][11].

본 논문에서는 비디오 전송 시스템에서 비디오 부호화기에 이용되는 양자화 파라미터(QP, Quantization Parameter)와 각 시간계층의 프레임 재생율 정보를 이용하여, 각 시간계층의 인코딩된 비트율과 화질을 예측하는 인공 신경망 모델을 제안한다.

제안된 비디오 전송 시스템은 급격히 변화하는 전송 대역폭의 변화에 대응하기 위해서, 스케일러블 비디오 부호화기를 이용한다. 그러나 전송 네트워크 대역의 변화에 대해서, 스케일러블 비트스트림을 생성하여 전송단에 저장할 수 없다. 네트워크 대역에 맞는 스케일러블 비트스트림이 송신단에 없다면, 비디오 부호화기를 이용하여 적당한 QP를 찾아서 부호화하고, 전송률과 대역폭을 비교하여 맞지 않으면, 다른 QP를 찾는 과정을 반복해야 할 것이다. 이러한 과정에서 수신단은 전송 지연으로 인한 낮은 QoS를 경험하게 된다.

우리는 인코딩된 스케일러블 비트스트림이 대응하지 못하는 전송 비트율에 대해서, 부호화기에 입력 파라미터만을 이용하여 전송 비트율과 화질을 예측할 수 있는 딥러닝 기반의 예측 모델을 제안한다. 제안된 모델은 다양한 비디오 부호화기의 매개변수의 영향을 이용하여 정확한 비트율 및 비디오 화질을 예측할 수 있으며, 이는 비디오 전송 시스템의 QoS 향상을 위해서 다양한 카메라 및 센서의 동적 설정 제어에 도움이 될 것을 기대된다.

논문의 구성은 다음과 같다. 2장에서는 제안된 비디오 전송 시스템에 대해서 설명한다. 3장에서는 다양한 모델을 개발하고 실험 설정과 모델링 방법론에 대해 설명한다. 4장에서는 검증 결과를 제시하고 전반적인 분석 결과를 보여준다. 마지막 섹션에서는 결론을 서술한다.

## II. 제안된 시스템 구성

그림 1은 비디오 전송 시스템(VTS, Video Transmission System)의 전체 구조도[1]를 보여준다. 송신단의 주요 기능은 수신단의 성능과 네트워크 대역폭에 맞추어 비디오 스트림을 끊임없이 전달하는 것이다. 이런 측면에서, 스케일러블 비디오는 입력 비디오를 여러 계층으로 분할하여, 개별적으로 부호화하기 때문에, 다양한 화질과 비트율을 갖는 비트 스트림을 생성하여 네트워크 대역폭 및 수신단의 성능에 따라 적절한 비디오 품질을 제공할 수 있다.

본 논문에서는 H.264/SVC 부호화기[2][3]를 이용하여, 입력되는 비디오 시퀀스를 시간 그리고 화질 측면의 스케일러블리티 특성을 갖도록 비트스트림을 생성한다. H.264/SVC 부호화기는 비디오 시퀀스와 이를 부호화하기 위한 다양한 파라미터를 입력 받는다. 양자화 파라미터(QP) 그리고 GOP (Group-of-Picture) 크기 등이 주요 스케일러블 비트 스트림 생성을 위한 파라미터이다. QP는 양자화 간격을 설정하며, QP가 낮으면, 양자화 간격이 좁아져, 사용되는 비트는 많아지고, 정보 손실은 적어진다. 반면 QP가 커지면, 양자화 간격이 넓어지고, 사용되는 비트수가 감소하며, 정보 손실이 많게 된다 [2]. QP는 0~51까지의 값의 범위에 존재하며, 네트워크 대역폭에 맞게 전송 비트율을 조정하기 위해

서, QP를 조정한다. 화질(Quality) 스케일러블리티는 비트율 조절을 위하여, 기저계층과 향상계층에 서로 다른 QP를 설정하여 부호화한다.

시간적(Temporal) 스케일러블리티는 1초간 재생되는 프레임수(fps, frame per seconds)를 조절하여, 전송 비트율을 조절한다. 그림 2는 GOP의 크기가 8인 경우, 계층적 예측구조(Hierarchical prediction structure)를 나타낸다[2]. 시간계층  $i$ 에서의 재생 프레임수를  $T(i)$ 로 정의한다. 입력 비디오 시퀀스가 30fps이면, 제일 높은 시간계층인  $T(4)$ 에서 30fps가 재생되고,  $T(1)$ 에서는 3.75fps가 재생된다. 네트워크 대역폭에 따라서 전송되는 프레임수와 비트율이 각각 조절된다

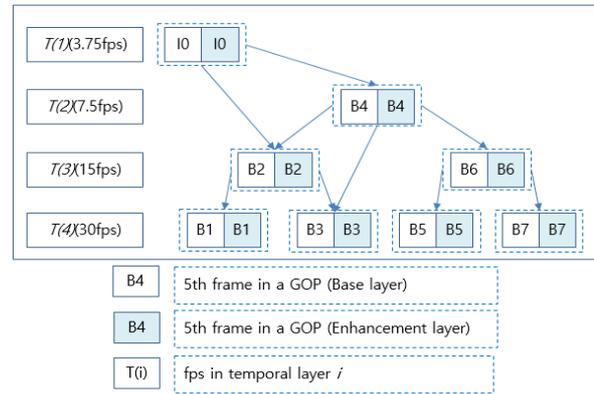


그림 2. GOP 크기가 8인 프레임에서의 계층적 예측 구조[2]

Fig. 2. Hierarchical prediction structure in 8 frame GOP size(30fps)[2]

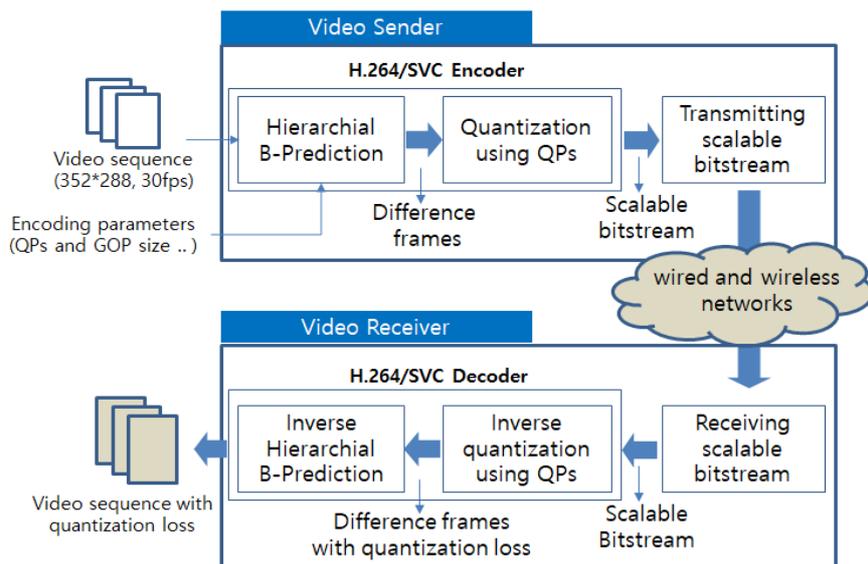


그림 1. 비디오 전송 시스템의 전체 구조도[1]

Fig. 1. Overall structure of video transmission system[1]

송신단은 생성된 비디오 비트스트림의 전송 비트율 및 화질을 알기 위해서는 직접 각각의 화질계층에 인코딩을 위한 QP를 설정하고, 비디오 시퀀스를 입력하여 부호화기 실행시켜야 하기 때문에, 비디오 전송 시스템의 전송 지연으로 인한 수신단의 QoS가 낮아질 수 있다.

본 논문에서는 비디오 전송 시스템에서 필요한 비디오 대역폭 및 인지되는 비디오 품질을 부호화 없이, 부호화기에 입력 정보만을 이용하여 예측할 수 있는 딥러닝 기반의 비트스트림 및 비디오 화질을 예측하는 신경망 모델을 개발한다. 제안된 모델은 다양한 비디오 부호화기의 매개변수의 영향을 이용하여 정확한 비트율 및 비디오 화질을 예측할 수 있으며, 이는 비디오 전송 시스템의 QoS 향상을 위해서 다양한 카메라 및 센서의 동적 설정 제어에 도움이 될 것으로 기대된다.

### III. 예측 모델 설계

2012년 이후 딥러닝(DL) 기술은 대용량 데이터의 이용, 높은 계산 능력, 그리고 강력한 인공 신경망(ANN) 알고리즘을 활용하여, 컴퓨터 비전 및 음성 인식, 자연어 처리, 모델링 등 다양한 분야에서 많이 이용되고 있다[10][11].

딥러닝을 위한 인공 신경망 알고리즘은 여러개의 층(layer)을 이용하여 인간의 뇌를 모방한다. 층은 하나의 입력층, 여러 개의 은닉층(Hidden layer) 그리고 한 개의 출력층으로 구성된다. 입력층은 하나의 독립 벡터로 한 개 이상의 입력 값들로 구성되며, 은닉 층은 하나 이상의 층으로, 인간 뇌의 신경계를 모델링한다. 그리고 출력층은 하나의 종속 벡터로 구성된다.

각 은닉층은 노드들로 구성되며, 해당 노드들은 다음 층의 모든 노드들과 연결된다. 각 층의 노드들은 이웃한 층의 가중치를 나타낸다. 가중치는 첫 번째 층이 두 번째 층에 어떻게 영향을 미치는지를 나타낸다. 훈련과정에서 출력 오차를 최소화하기 위해서 반복적으로 노드의 가중치가 업데이트 된다. 경사하강 역전파(Gradient descent back-propagation) 학습 알고리즘들이 이러한 출력 오차를 최소화하기 위해서 사용된다[9]. 각 노드에서의 최적의 가중치

를 계산하는 것은 딥러닝 및 머신러닝 훈련 과정의 주요 목표이며, 비선형성 특성을 갖는 가중치의 결정은 ReLU와 Softmax와 같은 각 노드의 활성화 함수(Activation function)를 이용하여 출력오차를 최소화 할 수 있다[9].

본 논문에서는 인공 신경망의 지도 학습(Supervised learning)을 활용한 딥러닝을 이용하여, 비디오 부호화기의 매개변수를 기반으로 비디오 화질과 비트율에 대한 예측모델을 제안한다. 레이블된 비디오 부호화기의 매개변수로 구성된 데이터를 기반으로, 높은 정확성을 갖는 예측 성능을 얻기 위해서, 인공 신경망의 회귀 분석(Regression analysis) 방식을 이용한다. 회귀분석은 어떤 변수가 다른 변수에 영향을 주고받는 경우, 영향을 주는 변수를 독립변수(Independent variable)라 하고, 영향을 받는 변수를 종속변수(Dependent variable)라고 한다. 회귀분석은 두 변수들간의 간의 관련성을 수식을 표현하고, 독립변수를 이용하여 종속변수의 변화를 예측(Prediction)하는 분석 방법이다[10].

그림 3은 딥러닝 과정을 보여준다. 비디오 시퀀스와 부호화 파라미터가 스케일러블 비디오 부호화기로 입력된다. 파라미터는 각 화질계층에서 사용될 QP값과 시간계층에서 사용될 fps 및 GOP 크기 등이 입력된다. 부호화 후에는 딥러닝 과정에서 사용될 각 시간계층별 비트스트림의 크기 및 비디오의 화질이 라벨링 된다. 비디오 화질 평가 메트릭으로 본 논문에서 PSNR(Peak-Signal-to-Noise-Ratio)(dB)를 사용한다. 라벨링된 데이터 값들은 ANN의 훈련 및 테스트를 위한 데이터셋(Dataset)으로 구분된다.

ANN 모델 구현을 위한 데이터셋을 만들기 위해서, 입력 독립변수는 각 화질계층에서 사용된 QP와 시간계층별 재생되는 프레임 수를 변화시키면서 비디오 시퀀스를 인코딩한 후, 그 결과로 출력되는 각 시간계층별 부호화된 비트스트림의 크기 및 비디오의 화질을 측정하여 출력 종속 변수로 설정한다. 훈련 및 테스트를 위해서 사용한 데이터셋은 'Mobile'과 'Bus' 비디오 시퀀스를 사용한다. 프레임수는 150개, 30fps이며, 프레임의 해상도는 352x288이다. GOP 크기는 16프레임으로 설정하여 5개의 시간계층 구성한다. 화질계층은 기저계층과 향상 계층을 각각 1개씩 갖는다.

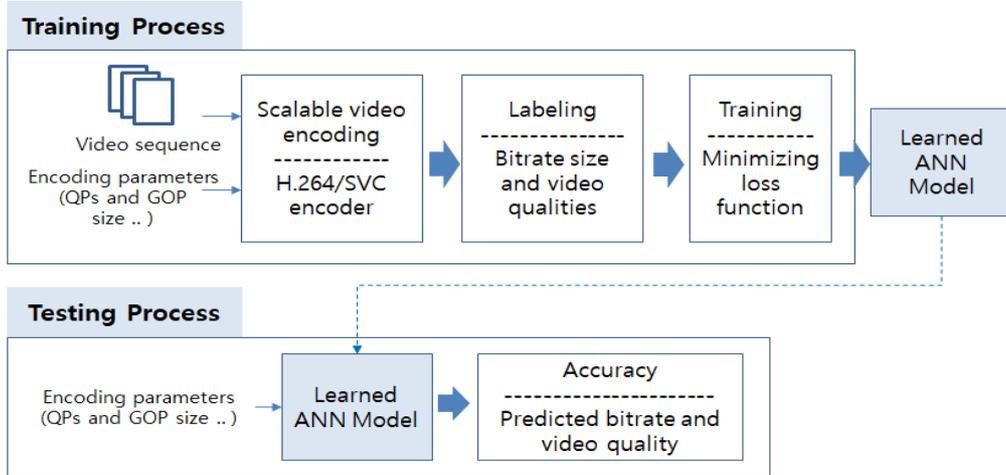
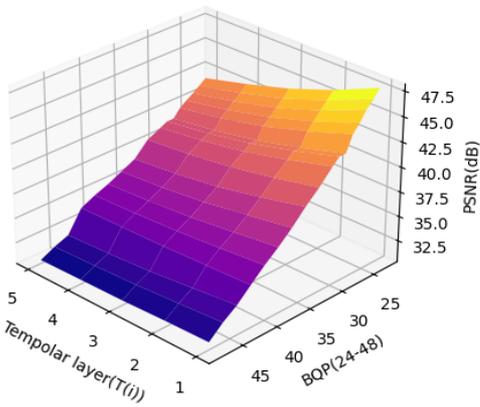
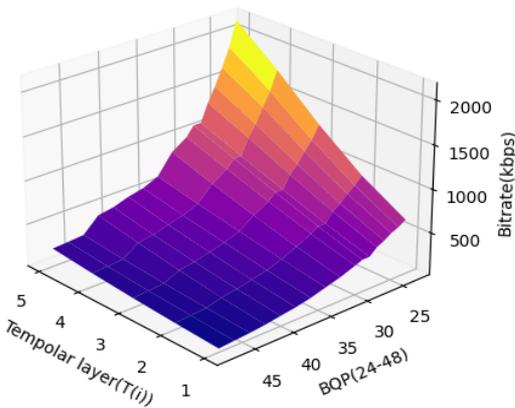


그림 3. 딥러닝 과정

Fig. 3. Description of deep learning process



(a) 화질(dB)  
(a) Image quality(dB)



(b) 비트율(Kbps)  
(b) Bitrate(Kbps)

그림 4. BQP와 시간계층(T(i))의 변화에 따른 비트율(kbps)과 화질(dB)의 관계

Fig. 4. Relation of bitrate and image quality according to variation of BQP and TL(i)

인코딩된 데이터를 이용하여 우리는 훈련과 테스트를 위한 데이터셋을 구성한다. 기저계층의 QP(Base quality layer QP, BQP)는 48,46,44,...,24로 48부터 2씩 감소하여 24까지 설정한다. 향상계층의 QP(Enhancement quality layer QP, EQP)는 BQP 보다 3 작은 값으로 설정한다. QP는 비디오 화질이 30~40dB 정도의 화질을 갖도록 설정하였다. EQP는 BQP보다 양자화 간격을 작게하여, 향상된 화질을 갖게 구성한다.

훈련단계에서는 훈련을 위해 준비된 데이터셋과 손실함수(Loss function)를 최소화하는 방향으로 학습이 진행된다. 훈련을 마친 학습된 모델은 학습에 사용되지 않은 테스트 데이터셋을 이용하여, 라벨링된 출력변수의 값을 예측한다. 예측된 데이터는 실제 데이터와 비교하여, 학습된 모델의 정확도(Accuracy)를 측정하며, 구현된 모델의 성능을 평가한다.

그림 4는 BQP와 각 시간계층 인덱스(i)에 대한 부호화된 비트스트림의 비트율과 비디오 화질의 관계를 각각 나타낸다. 비디오 화질 측면에서, 시간계층의 인덱스가 작아지고, QP가 작아질수록 PSNR 값이 선형적으로 증가하는 것을 알 수 있다. 이러한 경향은, 낮은 시간계층 인덱스에 포함되는 비디오 프레임은 계층적 예측구조 기반으로 참조되는 프레임 임수가 많기 때문에, SVC 부호화기는 낮은 시간 계층에 대해서 낮은 QP로 화질을 할당함으로써, 비트율이 증가하게 된다.

전송 비트율 측면에서, 시간계층의 인덱스(i)가 커지고, QP가 낮아질수록, 전송비트율은 비선형적으로 증가하는 것으로 볼 수 있다. 특히, QP가 30 이하로 낮아지고, 시간계층 인덱스가 4이상으로 커질 경우, 양자화 간격이 작아져서 전송 비트율이 급격히 증가하는 것을 볼 수 있다. 본 논문에서는 딥러닝 기반의 ANN 모델 설계 과정에서 Adam 옵티마이저와 활성화 함수로 ReLU를 사용함으로써, 이러한 비선형의 결과 예측을 수행할 수 있도록 모델을 설계한다.

본 논문에서는 파이썬 기반의 텐서플로우(Tensorflow)와 케라스(Keras)를 이용하여 ANN 모델을 구현한다[1,10]. 그림 5는 신경망 모델을 구현하기 위한 전체적인 프로세스를 보여준다. 우선 모델 구현을 위한 라이브러리 및 훈련 및 테스트를 위한 데이터를 구성한다. 세 번째 단계에서는 ANN 모델을 설계하기 위한 입력층, 출력층 그리고 은닉층의 수와 노드 수를 설정한다. 본 논문에서는 딥러닝을 위한 입력 독립변수로는 3개 컬럼 : BQP 와 EQP 그리고 T(i) (각 시간계층에서의 프레임 재생율)과 출력 종속변수 2개의 컬럼: 각 시간 계층별 인코딩된 비트스트림의 크기 및 화질로 구성된다. Adam 옵티마이저 및 ReLU 활성화 함수를 설정하여, 평균 제곱오차(MSE, Mean Square Error)를 최소화하도록 모델을 설계한다.

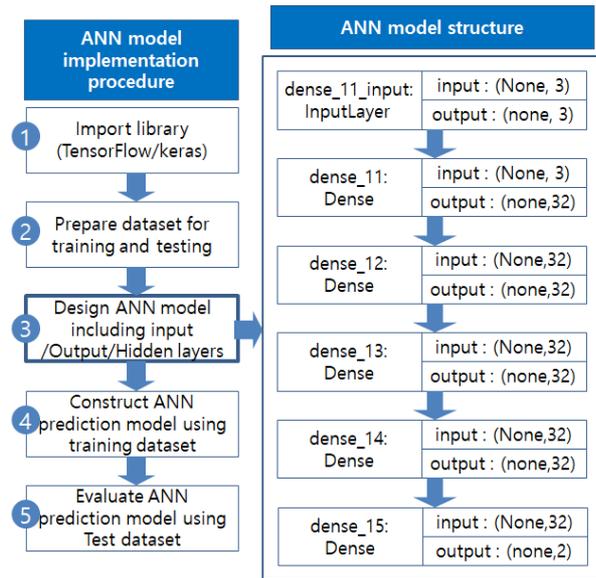


그림 5. 딥러닝 기반의 ANN 모델 구현을 위한 프로세스  
Fig. 5. Process of implementing ANN model based on deep learning

네 번째 단계에서는 훈련 데이터를 이용하여 모델 구현한다. 파이썬 객체인 model.fit()을 이용하여 진행되며, 에포크는 2000으로 설정한다. 마지막으로, 테스트 데이터셋을 이용하여 구현된 모델의 성능을 평가한다. 본 논문에서는  $R^2$ 를 사용하여 구현된 ANN 모델의 성능을 측정하며,  $R^2$ 가 1이면, 완벽한 적합(Perfect fit)을 나타낸다

#### IV. 모델 검증결과 및 분석

4장에서는 구현된 딥러닝 기반의 ANN 모델의 성능 검증과 분석을 제시한다. 그림 6는 에포크의 증가에 따른 제안된 신경망 모델의 출력 값인 비트율과 화질(PSNR)에 대한 검증 손실(Validation loss)의 결과를 나타낸다. 우리는 MSE를 손실함수로 사용한다.

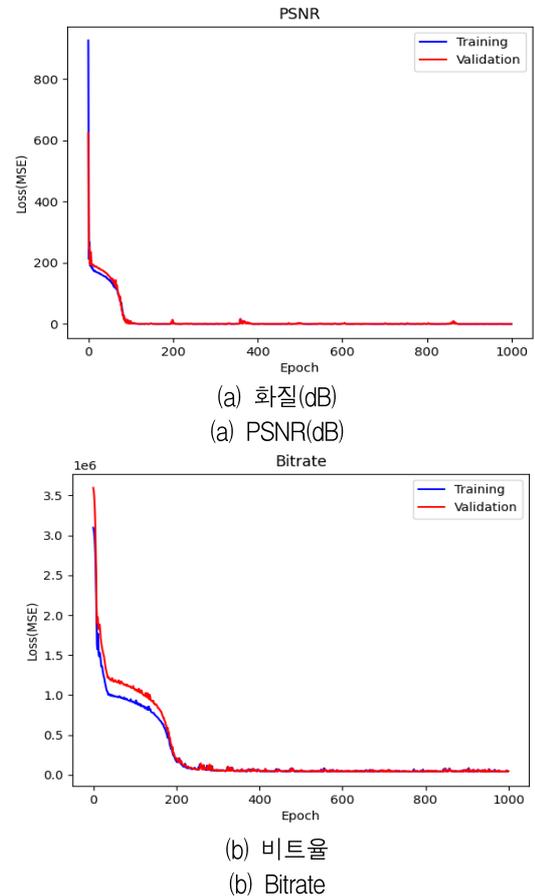


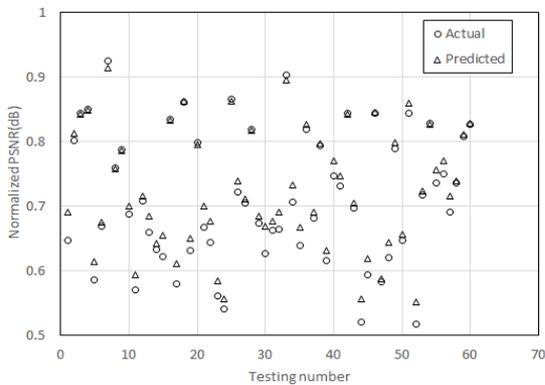
그림 6. 제안된 신경망 모델의 비트율과 화질(PSNR)에 대한 검증 손실 결과

Fig. 6. Results of validation loss in PSNR and Bitrate in proposed neural network model

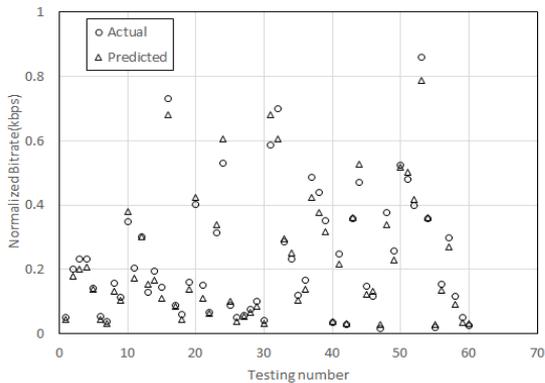
검증 손실을 측정하기 위해서, 훈련 데이터 셋에서 80%, 테스트 데이터셋에서 20%를 혼합하여 검증을 위한 데이터셋을 구성한다. MSE는 실제 데이터와 예측 데이터 사이의 차이를 제공한 후에 합을 평균하여 계산한다.

그림 6은 제안된 신경망 모델의 비트율과 화질(PSNR)에 대한 검증 손실 결과를 나타낸다. PSNR 예측의 경우, 100 에포크에서, 제안된 신경망 모델의 MSE가 0에 근접하는 것을 볼 수 있다. 반면, 비트율 예측의 경우, 값이 크기 때문에, 200 에포크에서 MSE가 0에 가까운 값으로 접근함을 볼 수 있다.

그림 7은 테스트 데이터셋에서 60개의 임의의 데이터를 추출하여 실제 데이터와 예측된 데이터를 보여준다. 비디오 화질(PSNR)의 예측 데이터는 실제 데이터와의 MSE가 작기 때문에 매우 근접한 출력 결과를 볼 수 있다.



(a) 최대신호대잡음비(dB)  
(a) PSNR(dB)



(b) 비트율  
(b) Bitrate

그림 7. 테스트 데이터셋에서 실제 데이터와 예측된 데이터의 비교

Fig. 7. Comparisons between actual and predicted data in the testing dataset (60 samples)

그림 4의 (a)에서 BQP와 EQP 및  $T(i)$ 의 기반의 PSNR 값은 선형적인 경향을 나타내었기 때문에 좋은 예측 결과를 볼 수 있다.

반면, 비트율의 경우, 높은 BQP에서는 선형적인 출력값을 나타내었지만, 낮은 BQP에서는 비트율이 급격하게 비선형적으로 증가함으로써, 큰 비트율에서는 예측값과 실제값에 약간의 차이가 있는 것을 확인할 수 있었다.

마지막으로 표 1은 제안된 ANN 모델의 성능을 나타내기 위해서  $R^2$  값을 나타내었다[1][9].  $R^2$  수치가 1에 가까워질수록 ANN 모델은 좋은 회귀 모델을 나타내며, 식 (1)은 다음과 같다.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (1)$$

여기에서,  $SS_{res}$ 는 실제값과 예측값간의 차이의 제곱의 합을 나타내며,  $SS_{tot}$ 는 실제값과 평균값간의 차이의 제곱의 합을 나타낸다. 비트율과 PSNR 값에 대한  $R^2$  결과값이 1에 가까운 값을 나타냄을 확인하였다. 전체적인 테스트 데이터에 대한  $R^2$  값은 0.98을 나타내었다.

표 1. 제안된 딥러닝 기반의 ANN 모델의  $R^2$  결과값  
Table 1.  $R^2$  results in proposed ANN prediction model

Data	$R^2$
PSNR(dB)	0.99
Bitrate	0.95
Overall testing	0.98

## V. 결 론

본 논문은 비디오 전송 시스템에서 딥러닝에 기반의 ANN 모델을 이용하여 비트율과 화질을 예측하는 모델을 제안하였다. 제안된 모델은 다양한 비디오 카메라 및 센서등에 적용되어, 현재 네트워크 대역폭에 기반하여, 예측된 인코딩 파라미터인 양자화 값과 스케일러블 비디오의 시간계층 인자를 설정함으로써, 비디오 화질 및 인코딩된 비디오의 비트율을 정확하게 조절할 수 있다. 제안된 ANN 모델은 예측된 비트율과 비디오 화질과 실제 데이터와 유사성 비교를 위해서,  $R^2$  값을 이용하였으며,

1에 가까운 값을 나타내어 정확도가 높은 예측 모델임을 확인하였다. 제안된 모델은 높은 해상도의 비디오 부호화기인 HEVC(High Efficient Video Coding) 등의 다른 비디오 부호화기에도 적용될 수 있을 것으로 기대된다.

### References

[1] A. Mohammad, Y. Sharrab, T. Monther, A. Sana, and S. Nabil, "Video coding deep learning-based modeling for long life video streaming over next generation", *Cluster Computing*, Vol. 26, pp. 1159-1167, Jan. 2023. <https://doi.org/10.1007/s10586-022-03948-x>.

[2] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 9, pp. 1103-1120, Sep. 2007. <https://doi.org/10.1109/TCSVT.2007.905532>.

[3] J. Reichel, H. Schwarz, M. Wien, and J. Vieron, "Joint scalable video model 9 of ISO/IEC 14496-10: scalable video coding", 2007.

[4] L. Wu and W. Zha., "Caching-Based Scalable Video Transmission Over Cellular Networks", *IEEE Communications Letters*, Vol. 20, No. 6, pp. 1156-1159, Jun. 2016. <https://doi.org/10.1109/LCOMM.2016.2555298>.

[5] Y. Ergiz, A. M. Dem, and T. Gir, "Joint multipath flow and layer allocation for scalable video streaming", *Computer Networks*, Vol. 191, May 2021.

[6] N. Kan, et al., "Deep Reinforcement Learning-based Rate Adaptation for Adaptive 360-Degree Video Streaming", *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, pp. 4030-4034, May 2019. <https://doi.org/10.1109/ICASSP.2019.8683779>.

[7] J. Luo, F. R. Yu, Q. Chen, and L. Tang, "Adaptive video streaming with edge caching and

video transcoding over software-defined mobile networks: A deep reinforcement learning approach", *IEEE Transactions on Wireless Communications*, Vol. 19, No. 3, pp. 1577-1592, Mar. 2020. <https://doi.org/10.1109/TWC.2019.2955129>.

[8] J. Saltarin, E. Bourtsoulatze, N. Thomos, and T. Braun, "Adaptive video streaming with network coding enabled named data networking", *IEEE Transactions on Multimedia*, Vol. 19, No. 10, pp. 2182-2196, Oct. 2017. <https://doi.org/10.1109/TMM.2017.2737950>.

[9] Y. Sharrab, I. Alsmadi, and N. Sarhan, "Towards the availability of video communication in artificial intelligence-based computer vision systems utilizing a multi-objective function", *Cluster Computing*, Vol. 25, pp. 231-247, Aug. 2021. <https://doi.org/10.1007/s10586-021-03391-4>.

[10] V. Mnih, et al., "Asynchronous methods for deep reinforcement learning", *Proc. of The 33rd International Conference on Machine Learning*, New York, USA, Vol. 48, pp. 1928-1937, Jun. 2016.

[11] M. Alsmirat, N. Al-Mnayyis, M. Al-Ayyoub, and A. Al-Mnayyis, "Deep Learning-Based Disk Herniation Computer Aided Diagnosis System From MRI Axial Scans", *IEEE Access*, Vol. 10, pp. 32315-32323, Mar. 2022. <https://doi.org/10.1109/ACCESS.2022.3158682>.

### 저자소개

하 호 진 (Hojin Ha)



1999년 2월 : 명지대학교  
 제어계측공학과(공학사)  
 2000년 2월 : 한양대학교  
 제어계측공학과(공학석사)  
 2009년 2월 : 연세대학교  
 전기전자공학부(공학박사)  
 2000년 2월 ~ 2011년 8월 :

삼성전자 DMC 연구소 책임연구원

2018년 3월 ~ 현재 : 한양여자대학교 소프트웨어융합과  
 교수

관심분야 : 멀티미디어 신호처리 및 네트워킹