Journal of Industrial Information Technology and Application *ISSN(Online):2586-0852*
Vol.7 No.1

A Deep Learning based Audio Super-Resolution Algorithm from Lossless and Lossy Input Data

Hong-Jin Kim [1], Jeong Tak Ryu[2] and Kyuman Jeong[3*1]
[1] *School of AI, Daegu University, Daegu, Korea*
[2] *College of Information and Communication Engineering, Daegu University, Daegu, Korea*
[3] *School of AI, Daegu University, Daegu, Korea*

Abstract. Artificial intelligence technology, in which computers perform human-like actions or behaviors, is becoming popular. Particularly, efforts are being made to implement technologies that classify objects or respond to user behavior. It is also attracting attention in fields that require much time and effort, such as restoring paintings drawn in the past. It is expected that it can be used in various fields as well as an image restoration technique using three-dimensional data. In particular, audio data has changed from the way of using physical storage devices in the past to the way of being provided on a network basis. In this paper, we propose an algorithm to recover high - quality audio data from the internal storage device that can be self - produced by receiving compressed audio data. We propose a method of restoring audio data that is arranged and reproduced by changing time-dependent one-dimensional data using lossless audio data and lost audio data after compression through a deep learning technology, CNN (Convolutional Neural Network).

Keywords; component, formatting, style, styling, insert

* Corresponding author e-mail: kyuman.jeong@gmail.com
Received: Jan 16, 2023; Accepted: Mar 1, 2023; Published: Mar 31, 2023

[1] This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022S1A5C2A07091326)

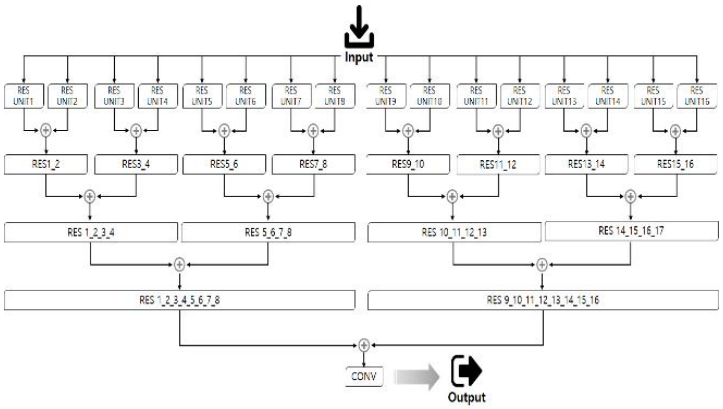


Fig. 1. Overall Process

1. Introduction

Recording information contributes to the greatest contribution to the advancement of human technology. In the past, it was common to pass on to future generations by recording texts or pictures, but in modern times, sound and video can be recorded, and various information can be easily accessed. In particular, the act of recording sound rather than video has been developed so that it can be easily transmitted and miniaturized from phonographs to LPs, cassette tapes, CDs, and MP3s. The biggest reason that we were able to move from the material realm to the intangible is the recording method in a file format that can be easily transmitted. With the rapid rise of the Internet and smartphone penetration, the way data was used in the past using physical storage devices has changed. In particular, more data than a storage medium that a machine can hold, such as a network-based cloud service, can be provided through the Internet, and streaming services that receive video or sound data wirelessly are in the spotlight. In this paper, to solve the size and time constraints in Internet-based data transmission, we propose a technology that reduces the size of transmitted data and enables efficient audio streaming through self-restoration in the mechanical device responsible for physical storage and execution.

2. Proposed Method

In this paper, we study how to convert low-quality audio data into high-quality data using CNN used for media restoration. In the process of restoring sound quality using deep learning, WAV format data extracted from CD is converted into low-quality data and high-quality data, and used as input values and GT (GroundTruth) to be used for

learning, respectively. Repeated learning We want to design an algorithm that restores audio data through The basis of the model used for learning is to compare and evaluate various models using CNN to select the model with the highest degree of completeness, and to introduce the process of designing a new model by modulating the model if necessary.

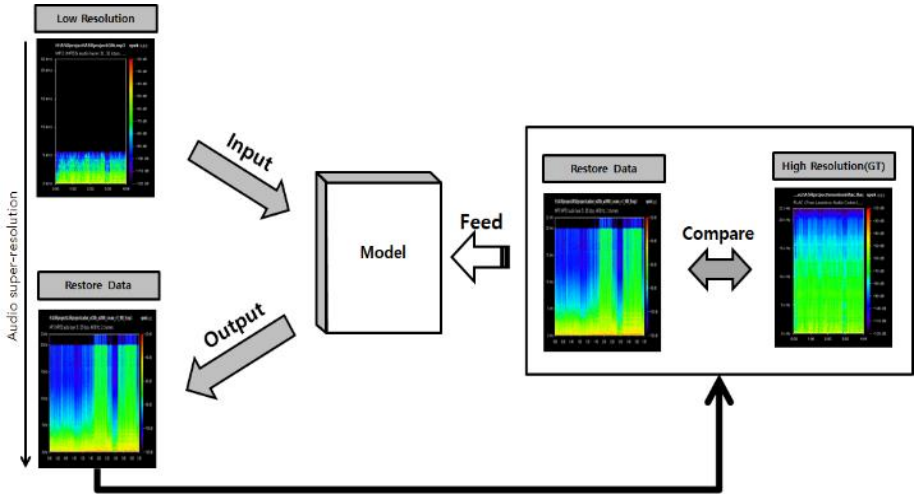


Fig. 2. Audio Quality Restoration Algorithm

A. Data set generation

In the learning data generation process, 101 sound source CDs are used, and since each audio file has a different playback time, it is necessary to establish a unified standard. The data generation process is a total of two steps, extracts the audio data stored in the first CD, converts it into a lossless compression format Flac, which is the standard of the restoration point, and saves it. Next, in order to generate low-bitrate audio data to be used as input data, the Flac file extracted from the CD is down-sampled and generated. The number of extracted sound sources is a total of 1,137, and the compressed data must be decoded using the Pydub library that supports FFmpeg (Febrice Bellard et). Convert to format.

B. Learning model

In this paper, we implement and implement a method to restore audio quality through extension by controlling the audio bandwidth using deep learning. The methods used for implementation were DRRN (Deep Recursive Residual Network) and SRResNet (Super-Resolution Residual Network) models used in Image Super-Resolution. Unlike

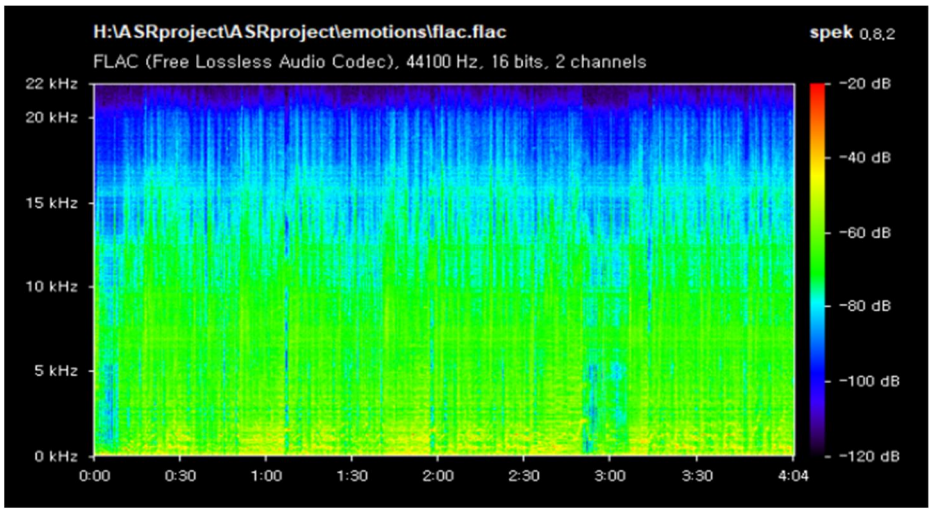
image data, audio data stored in a one-dimensional array is used. Therefore, a new model is designed and proposed to have higher long-term accuracy, and it undergoes a process of comparative analysis with various previously presented CNN models

Compared with ResNet and DRRN, our proposed model is shown in Figure 1, and RES_UNIT uses the Residual Unit structure of the DRRN algorithm.

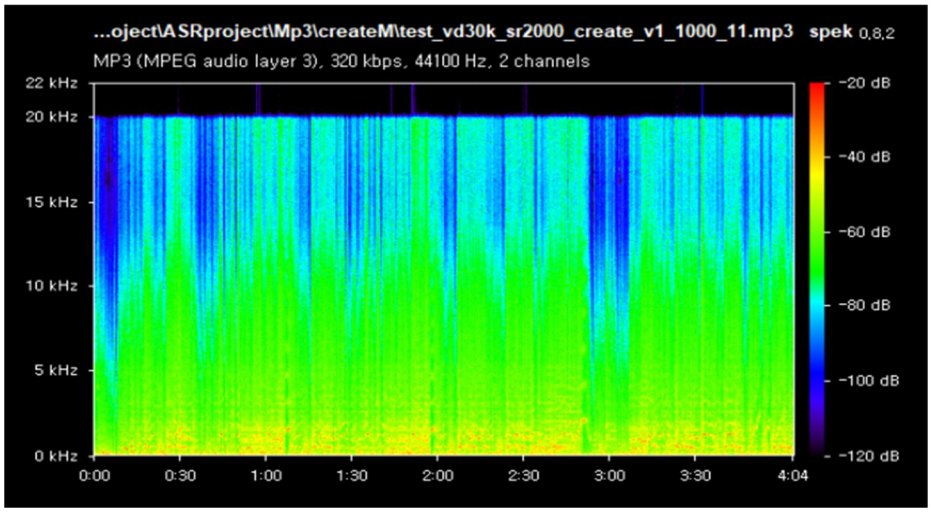
For the initial input, not a single residual unit, but multiple residual units are used to extract features or patterns of different data, and convolution is performed to the next layer by adding two equation pairs to the passed results, and the last one is passed through the residual unit. One result is passed through a single convolution map to derive the output.

C. Experimental results

After dividing the total 1,137 raw data by channel, it was studied using 4,512,753 datasets of High Resolution data and Low Resolution data, respectively, prepared by dividing into 2,000 sizes. The spectrum of the sound source to be used as the test set after learning is shown in Figure 3 below.



(a) Original sound



(b) Learning result

Fig. 3. Spectrum for the original sound source (a) and the result obtained through training (b)

3. Conclusions

As models to be used in the training process, ResNet and DRRN, which are CNN models that are often used to restore high-resolution images from low-resolution, were used. When restoring the sound quality, it was confirmed that the model used in the existing Image Super-Resolution could not obtain good performance. As shown in Figure 34, although the high-pitched data is alive, there is a problem that it does not match the original spectrum. which is still different from In addition, noise generated during the restoration process can be identified. The time until the sound source derives the result through the model also takes about 3 to 5 minutes when using the current 3 minute sound source. Since audio has a lower data dimension than images, the proposed model uses a number of Residual Units at the initial input rather than the size or number of filters or the depth of the layer, and learns in pairs. It has been confirmed, and it is expected that better results can be obtained if the number of training times, the size and number of filters, and the size of the data to be learned are adjusted.

References

[1] Ying Tai, Jian Yang and Xiaoming Liu. (2017). Image super-resolution via deep recursive residual network. In CVPR

[2] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee. (2016). Accurate image super-resolution using very deep convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition3.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. (2016). Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition.

[4] V. H. Quintana and Edward J. Davison. (1974). Clipping-off gradient algorithms to compute optimal controls with constrained magnitude. International Journal of Control, vol. 20, no. 2, pp. 243-255.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. (2012). Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems.

[6] Jiwon Kim, Jung Kwon Lee and Kyoung Mu Lee. (2016). Deeply-recursive convolutional network for image super-resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[7] Joan Bruna, Pablo Sprechmann and Yann LeCun. (2016). Super-resolution with deep convolutional sufficient statistics. In International Conference on Learning Representations (ICLR).

[8] Volodymyr Kuleshov, S. Zayd Enam and Stefano Ermon. (2017). Audio super-resolution using neural nets. Presented at the 5th International Conference on Learning Representations (ICLR)