Edge Storage Management Recipe with Zero-Shot Data Compression for Road Anomaly Detection

YeongHyeon Park
SK Planet Co., Ltd.
Seongnam, Rep. of Korea
yeonghyeon@sk.com

Uju Gim
SK Planet Co., Ltd.
Seongnam, Rep. of Korea
gim.uju1217@sk.com

Myung Jin Kim
SK Planet Co., Ltd.
Seongnam, Rep. of Korea
myungjin@sk.com

***Abstract*—Recent studies show edge computing-based road anomaly detection systems which may also conduct data collection simultaneously. However, the edge computers will have small data storage but we need to store the collected audio samples for a long time in order to update existing models or develop a novel method. Therefore, we should consider an approach for efficient storage management methods while preserving high-fidelity audio. A hardware-perspective approach, such as using a low-resolution microphone, is an intuitive way to reduce file size but is not recommended because it fundamentally cuts off high-frequency components. On the other hand, a computational file compression approach that encodes collected high-resolution audio into a compact code should be recommended because it also provides a corresponding decoding method. Motivated by this, we propose a way of simple yet effective pre-trained autoencoder-based data compression method. The pre-trained autoencoder is trained for the purpose of audio super-resolution so it can be utilized to encode or decode any arbitrary sampling rate. Moreover, it will reduce the communication cost for data transmission from the edge to the central server. Via the comparative experiments, we confirm that the zero-shot audio compression and decompression highly preserve anomaly detection performance while enhancing storage and transmission efficiency.**

***Index Terms*—anomaly detection, data compression, edge computing, storage management, transmission efficiency**

I. Introduction

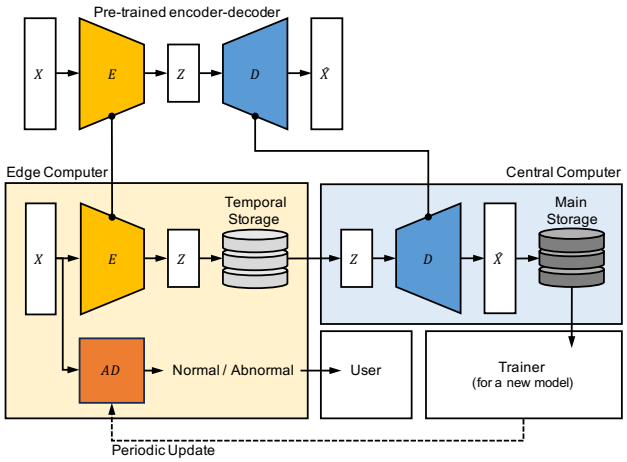
Knowing road conditions is an effective way to prevent traffic accidents [1]. Most of the road hazards are highly related to icy or wet roads which reduce the friction between the road and tires. When considering a vision sensor-based road anomaly detection system [2]–[5], inclement weather will make occlusion on the camera which makes it difficult for understanding road conditions [6], [7]. Moreover, intensity-changing situations such as at night also make it difficult to determine road conditions.

As an approach to solving these problems, an audio-based anomaly detection approach has been developed. The audio-based system receives information from the medium wave in the air, so it shows a better response-ability than the occlusion situation of the vision sensor. In addition, sound can be properly transmitted even at night time, an audio-based system is recommended for this situation.

Even in the case of successful anomaly detection as above, continuous updating of the anomaly detection model is required considering that target environments are continuously aging [8]. For this, we need high-quality large data to update the neural network-based anomaly detection model. We have installed edge computers on the road for anomaly detection, which has limited resources such as storage. Considering the above, we have a limitation to keep audio data for the long term. The above limitation can be partially mitigated

(a) Edge computer

(b) Road sign for hazard warning



(c) Pre-trained encoder-decoder-based storage management system

Fig. 1. Proposed scheme of edge storage management system for road anomaly detection. An edge computer (a) that has a small storage space is installed on the habitual freezing section (b). A proposed method for saving as much data as possible in the limited storage space is shown in (c). Among the pre-trained encoder-decoder, the encoder and decoder are deployed on the edge and central computer respectively. Encoder on edge computer encodes the collected high-resolution audio into latent code Z and the saved Zs are transmitted to the central server at regular intervals. The central server decodes the received latent codes for original high-resolution audio.

by transmitting the audio to large central storage in time, but in this case, enormous data transmission costs will be incurred [9]. Also, we can lower the quality of the audio collected from high fidelity (Hi-Fi) to low fidelity (Lo-Fi) to reduce the file size for each audio, but this is not recommended as it leads to fundamental information loss.

Motivated by this, we propose a storage management method that can keep as much data as possible in the edge computer for a long time in limited storage space while minimizing the cost of the data transmission into a central server. Our method is based on zero-shot encoding and decoding using a pre-trained audio super-resolution (ASR) model [10]–[12]. Referring ASR models can convert input audio to high-resolution, so they can handle arbitrary resolution inputs that are lower than the target resolution they are trained on.

The overall scheme of our proposal is shown in Fig 1. The edge computer continuously collects data and determines whether the situation is abnormal or not through a microphone facing the road. Basically, collected audio is stored in the original resolution, but it is saved after being converted into a latent vector by an encoder of pre-trained ASR. The latent vectors, stored in edge storage, are sent to the central server at regular intervals or the storage space fills up to a certain level. Then, at the central computer, they are restored to audio form by a decoder paired with the above encoder [9]. At this time, even if the resolution of the collected sound sources is different, it is characterized by being restored to a similar Hi-Fi quality by the decoder of the central server. The restored audio data is used for the purpose of developing a novel anomaly detection model or updating existing models.

To verify the proposed method, we collected data from three roads with different characteristics. The audio is basically collected at 44,100 Hz. For the purpose of saving storage space experiments in which downsampling is applied to assume a situation in which a microphone such as 11,025 Hz, and our zero-shot audio encoding method are also covered.

Overall, our contributions are summarized below:

- When downsampling is applied, high-frequency information loss occurs, confirming that hinders precise anomaly detection. Our approach, zero-shot encoding and decoding minimize information loss and preserves anomaly detection performance at an appropriate level while maximizing storage efficiency.
- We show that there is no need to train new models to encode and decode for our domain, road noise. Our example deals with road anomaly detection as a target, but our method can also be extended to another edge computer-based data collection approaches in other domains.

II. Related works

A. Road condition identification

To identify abnormal situations on the road such as road bumps, cracks, or potholes, methods based on vision sensors have been proposed [2]–[5]. However, since the abnormal situation on the road is highly related to bad weather, and bad weather can obscure the view of the camera [6], [7], a vision sensor-based approach will show constrained detection performance. Some approaches using motion or gyroscope sensors rather than a vision sensor can partially ease the above problem but it shows unstable performance that highly depends on pre-defined settings [13], [14]. An audio-based anomaly detection method has been proposed as a way to overcome the problem of invisible situations to make decisions in bad weather or night situations [1], [15].

(a) Tunnel (b) City (c) Outer

Fig. 2. Three data collection sites. Each has different road characteristics. The tunnel, shown in (a) has sound reverberation, and the city, (b), has irregular reflection by facilities. In the case of the outer road, (c), there is no sound reflection almost.

To construct a reliable road anomaly detection system in outdoor environments, we inherit the above approach from prior research to take advantage of audio-based road anomaly detection.

B. Data compression

The most intuitive way to maximize data storage efficiency is to collect data at a lower resolution. However, when we need to create a high-quality anomaly detection model, we also need high-resolution data [16].

The computational approach which collects high-resolution data and encodes it into lower dimensions can be considered other than the above [9], [17]. When the encoding method is provided with a paired decoding method, we can easily compress the data into small sizes and decompress them into the original resolution. In this case, some information loss may occur in the process of encoding and decoding, but it is recommended as an alternative to hardware-based capacity-saving methods that block information fundamentally.

An artificial neural network-based encoder-decoder (ED) shows better reconstruction performance than traditional methods [18], [19]. In particular, a model trained for super-resolution (SR) purposes is useful as a method to help roughly estimate the high-frequency region of data collected at lower resolution [10]–[12].

We propose a storage space management method based on zero-shot encoding decoding using the pre-trained ED for SR purposes, considering that the resolution of audio sensors installed in each region may be different.

III. Approach

A. Overview

The overall of our proposal is shown in Fig. 1. which is an encoding and decoding system for storing as much audio

data as possible in an edge computer. We separate pre-trained ED into each component encoder and decoder. Then, we locate each of the above on the edge and central computers respectively. Our approach only uses a single encoder and decoder pair rather than having each model for each different road environment as shown in Fig. 2. This is dubbed as a zero-shot inference that can eliminate the hassle of preparing models to reflect the surrounding environment of countless sensors installed at outdoor points.

B. Audio super-resolution

For data compression, we utilize a pre-trained ASR model, EnCodec [12], which is constructed with an encoder and decoder. The sensors installed at each site to perform the road anomaly detection that we will cover may include expensive high-resolution microphones or low-resolution microphones, depending on the management budget of the local government. The advantage of using the ASR model is that the input data can be converted into high-resolution data regardless of the input resolution. This allows audio data collected from microphones of arbitrary resolutions as aforementioned can be integrated into Hi-Fi audio.

Any ASR model can be employed for the encoding and decoding process, but a structure in which the encoder and decoder can be used separately is recommended. A method of performing information augmentation in a feature map or latent vector stage may rather increase the capacity of encoded data [10], so it should be avoided.

C. Zero-shot encoding and decoding

It is difficult to obtain a pre-trained ASR model on the road noise domain because the audio corresponding to the friction noise between the tire and the road surface, which we deal with, is not commonly used data. In addition, the number of sensors installed on the roads we deal with is numerous and their characteristics are highly diverse, so it takes a huge amount of time and cost to build a model while guaranteeing the generalization ability.

As a way to easily overcome these methods, we adopt a zero-shot inference that utilizes a highly generalized ASR model which pre-learned with a wide range of audio data including general audio, speech, and music. Following the above, we propose a method to separate the encoder of the pre-trained ASR model, place it on the edge computer and encode all the collected data. The encoded data will be transmitted o the central computer and decoded.

IV. Experiments

A. Dataset

To validate the zero-shot inference-based method, we should deal with varied data from different road environments. Among the collectible road points, we selectively use three points with significantly different environmental characteristics as shown in Fig. 2. We have collected the audio samples for four weather conditions at three locations as summarized in Table I.

TABLE I
SUMMARY OF THE DATASET ACQUIRED FROM THE THREE DIFFERENT LOCATIONS, IN NUMBER OF AUDIO (NUMBER OF DRIVING EVENTS) FORM. EACH AUDIO IS A 10-MINUTE LENGTH AND 10-SECOND LENGTH DRIVING EVENTS ARE EXTRACTED FROM THE AUDIO.

Post	**Normal**		**Abnormal**					
	Dry		**Wet**		**Slush**		**Snow**	
Tunnel	10	(384)	10	(21)	4	(7)	-	-
City	10	(804)	10	(529)	2	(11)	-	-
Outer	10	(1,153)	9	(1,032)	10	(76)	3	(5)
Total	30	(2,341)	29	(1,582)	16	(94)	3	(5)

TABLE II
MEASURED FILE SIZE OF ONE-SECOND LENGTH AUDIO DATA FOR EACH FILE MANAGEMENT SETTING

Source	**Hi-Fi**	**Lo-Fi**		**ASR**
Frequency	f_{44} **[15]**	f_{22}	f_{11}	f_{44}
File size	173 KiB	87 KiB	44 KiB	5 KiB
Ratio$_{size}$ $\downarrow$	1.000	0.503	0.254	**0.029**

B. Zero-shot compression

Referring to our purpose, maximizing data compression, it is important to minimize the restoration error as well as the compression capacity of the data. Note that, we abbreviate 44,100 Hz, 22,050 Hz, and 11,025 Hz as f_{44}, f_{22}, and f_{11} respectively. The f_{22} and f_{11} in Fig. 3 show the fundamental loss of high-frequency components compared to f_{44} which the sampling rate is set as low to reduce the file size. Therefore, it should be avoided to set the low sampling rate with a hardware-based approach, and a method of collecting and compressing Hi-Fi data needs to be used in a computational approach.

When applying a pre-trained ASR model for audio compression and decompression, it shows not only preserving high-frequency components but also less information loss than the hardware-based approach as shown in $\hat{f}_{44}$ in Fig. 3 and Table II.

Through this experiment, we confirm that the data compression method based on the ASR model can increase the total amount of samples saved in an edge computer by $34.6\times$ while minimizing information loss. This means that the cost of data transmission can also be $34.6\times$ reduced.

C. Anomaly detection

We simulate the compressed data collecting situation by the zero-shot encoding method in the central computer to check whether an anomaly detection model can perform at the appropriate level when it is trained with the decompressed dataset. It is clear that zero-shot encoding is a method that can minimize information loss while maximizing compression rate compared to others, but considering that there is a slight difference from the original, we can estimate that anomaly detection performance may also be decreased. Considering this, the method with the least performance degradation can be considered as the optimal method.

For the experiments, we downsample each audio by 2 and 4 scales to simulate the same Hi-Fi data as collected at Lo-Fi

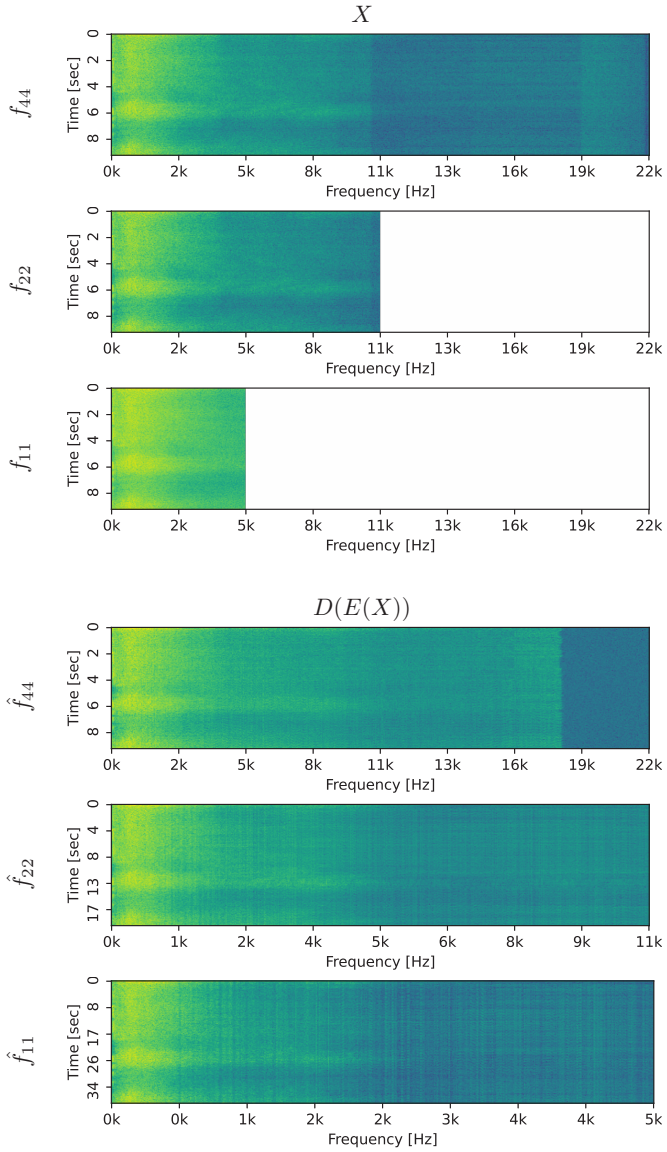


Fig. 3. Results of encoding and decoding for each audio input X with EnCodec [12]. Each encoding and decoding is abbreviated as E and D.

conditions. Note that, the resolution of the original audio is 44,100 Hz, resolution for each downsampled audio is 22,050 Hz and 11,025 Hz respectively. Also, the compressed and decompressed audio data are used to verify the ASR case. The measured anomaly detection performance with the area under the receiver operating characteristic curve (AUROC) [20] is summarized in Table III. In the case of using ASR, the average performance decreased to 92%-level compared to the original Hi-Fi audio case. However, we confirm that the anomaly detection performance of our method is more compliant than Lo-Fi.

D. Resolution integration

We verify that it can be integrated into equal-level of high-resolution audio through the encoding and decoding process when the resolution (sampling rate) of the collected audio is different. If this premise is satisfied, it can guarantee that the proposed data compression and restoration framework via the ASR model works properly no matter which audio resolution.

If the data collected at low-resolution can be converted into high-resolution, it can be helpful to build a high-performance anomaly detection model considering the case where low-cost and low-resolution microphones are inevitably installed according to the budget. When the three samples, assuming original high-resolution data and low-resolution data, are up-sampled through the ASR model, they show almost the same difference from the original as summarized in Table IV. Thus, we confirm that data of arbitrary resolution can be integrated into high-quality audio at the central server.

TABLE III
ANOMALY DETECTION PERFORMANCE AT EACH AUDIO FREQUENCY.

Source	**Hi-Fi**	**Lo-Fi**		**ASR**
Frequency	f_{44} **[15]**	f_{22}	f_{11}	f_{44}
Tunnel	0.963	0.961	0.957	**0.967**
City	0.871	0.752	0.794	**0.831**
Outer	1.000	0.841	0.818	**0.847**
Merge	0.915	0.803	0.791	**0.812**
Average	0.937	0.839	0.840	**0.864**
Ratio$_{AUROC}$ ↑	1.000	0.895	0.896	**0.922**

TABLE IV
RECONSTRUCTION ERROR, MEAN SQUARED ERROR (MSE), FOR EACH SOURCE AUDIO FREQUENCY.

Frequency	f_{44}	f_{22}	f_{11}
MSE$(f, \hat{f})$	71.07297	71.07244	71.07209

V. CONCLUSION

We propose a method based on the pre-trained ASR model for storing as many audio samples as possible in an edge computer with limited storage capacity installed for the purpose of road anomaly detection. Our method shows that adequate performance can be obtained by using only one generalized encoder-decoder pair instead of each encoder-decoder corresponding to each post or type of road with high environmental diversity. Each audio is highly compressed from its original size of 173 KiB per second to 5 KiB, showing that it can store up to $34.6\times$ as many audio samples. In addition, even if an anomaly detection model is trained by collecting compressed audio samples at the central computer, an appropriate level can be achieved. Some degradation of the anomaly detection performance is caused by a slight information loss during the encoding and decoding process but there is room for improvement via proper encoder-decoder pairs. In future work, we plan to explore the encoder-decoder model with better generalization ability or trained on the road noise domain as a way to construct more stable systems.

ACKNOWLEDGEMENTS

We are grateful to all the members of SK Planet Co., Ltd., who have supported this research, providing equipment for the experiment.

REFERENCES

[1] YeongHyeon Park, and JongHee Jung. "Efficient Non-Compression Auto-Encoder for Driving Noise-based Road Surface Anomaly Detection." *IEEJ Transactions on Electrical and Electronic Engineering* (2022).

[2] Seung-Ki Ryu, Taehyeong Kim, and Young-Ro Kim. "Image-based pothole detection system for ITS service and road management system." *Mathematical Problems in Engineering* 2015 (2015): 1-10.

[3] Rui Fan, Mohammud Junaid Bocus, Yilong Zhu, Jianhao Jiao, Li Wang, Fulong Ma, Shanshan Cheng, and Ming Liu. "Road crack detection using deep convolutional neural network and adaptive thresholding." 2019 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2019.

[4] Rozi Bibi, Yousaf Saeed, Asim Zeb, Taher M Ghazal, Taj Rahman, Raed A Said, Sagheer Abbas, Munir Ahmad, and Muhammad Adnan Khan. "Edge AI-based automated detection and classification of road anomalies in VANET using deep learning." *Computational intelligence and neuroscience* 2021 (2021): 1-16.

[5] Tomas Vojir, Tomáš Šipka, Rahaf Aljundi, Nikolay Chumerin, Daniel Olmeda Reino, and Jiri Matas. "Road anomaly detection by partial image reconstruction with segmentation coupling." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021.

[6] Nastaran Yaghoobi Ershadi and José Manuel Menéndez. "Vehicle tracking and counting system in dusty weather with vibrating camera conditions." *Journal of Sensors* 2017 (2017).

[7] Kun Qian, Shilin Zhu, Xinyu Zhang, and Li Erran Li. "Robust multimodal vehicle detection in foggy weather using complementary lidar and radar signals." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021.

[8] Young Jong Song, Ki Hyun Nam, and Il Dong Yun. "Anomaly Detection through Grouping of SMD Machine Sounds Using Hierarchical Clustering." *Applied Sciences* 13.13 (2023): 7569.

[9] Joseph Azar, Abdallah Makhoul, Mahmoud Barhamgi, and Raphaël Couturier. "An energy efficient IoT data compression approach for edge machine learning." *Future Generation Computer Systems* 96 (2019): 168-175.

[10] Yuang Li, Yuntao Wang, Xin Liu, Yuanchun Shi, Shwetak Patel, and Shao-Fu Shih. "Enabling Real-Time On-Chip Audio Super Resolution for Bone-Conduction Microphones." *Sensors* 23.1 (2022): 35.

[11] Seungu Han, and Junhyeok Lee. "NU-Wave 2: A general neural audio upsampling model for various sampling rates." arXiv preprint arXiv:2206.08545 (2022).

[12] Alexandre Défossez, Jade Copet, Gabriel Synnaeve, and Yossi Adi. "High fidelity neural audio compression." arXiv preprint arXiv:2210.13438 (2022).

[13] H. Bello-Salau, A.M. Aibinu, A.J. Onumanyi, E.N. Onwuka, J.J. Dukiya, and H. Ohize. "New road anomaly detection and characterization algorithm for autonomous vehicles." *Applied Computing and Informatics* 16.1/2 (2018): 223-239.

[14] Shahram Sattar, Songnian Li, and Michael Chapman. "Developing a near real-time road surface anomaly detection approach for road surface monitoring." *Measurement* 185 (2021): 109990.

[15] YeongHyeon Park, Myung Jin Kim, and Won Seok Park. "Frequency of Interest-based Noise Attenuation Method to Improve Anomaly Detection Performance." 2023 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2023.

[16] Yuxiang Zhao, Wenhao Wu, Yue He, Yingying Li, Xiao Tan, and Shifeng Chen. "Good practices and a strong baseline for traffic anomaly detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021.

[17] Shaofei Lu, Qinhua Xia, Xiaolin Tang, Xuyang Zhang, Yingping Lu, and Jingke She. "A reliable data compression scheme in sensor-cloud systems based on edge computing." *IEEE Access* 9 (2021): 49007-49015.

[18] Shifeng Zhang, Ning Kang, Tom Ryder, Zhenguo Li. "iflow: Numerically invertible flows for efficient lossless compression via a uniform coder." Advances in Neural Information Processing Systems (NeurIPS) 34 (2021): 5822-5833.

[19] Joseph Azar, Gaby Bou Tayeh, Abdallah Makhoul, and Raphaël Couturier. "Efficient Lossy Compression for IoT Using SZ and Reconstruction with 1D U-Net." *Mobile Networks and Applications* 27.3 (2022): 984-996.

[20] Fawcett, Tom. "An introduction to ROC analysis." *Pattern recognition letters* 27.8 (2006): 861-874.