IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 April 2024, accepted 30 April 2024, date of publication 6 May 2024, date of current version 17 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3397842

RESEARCH ARTICLE

Disentangled Prototypical Convolutional Network for Few-Shot Learning in In-Vehicle Noise Classification

ROBIN INHO KEE[1], DAHYUN NAM[2], SEOK-JUN BUU[3], (Member, IEEE), AND SUNG-BAE CHO[4], (Senior Member, IEEE)

[1]Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, USA
[2]Department of Mechanical Engineering, Yonsei University, Seoul 03722, Republic of Korea
[3]Department of Computer Science, Gyeongsang National University, Jinju, Gyeongsangnam-do 52828, Republic of Korea
[4]Department of Computer Science, Yonsei University, Seoul 03722, Republic of Korea

Corresponding authors: Seok-Jun Buu (sj.buu@gnu.ac.kr) and Sung-Bae Cho (sbcho@yonsei.ac.kr)

This work was supported in part by the Yonsei University Research Fund (Postdoctoral Researcher Supporting Program) of 2023 under Grant 2023-12-0021, in part by the Yonsei Fellow Program funded by Lee Youn Jae, in part by Institute for Information and Communication Technology Planning and Evaluation (IITP) Grant funded by Korea Government [Ministry of Science and ICT (MSIT)] (Artificial Intelligence Graduate School Program, Yonsei University) under Grant 2020-0-01361, and in part by the Air Force Defense Research Sciences Program funded by the Air Force Office of Scientific Research.

ABSTRACT This study addresses the persistent challenge of in-vehicle noise, a significant factor affecting customer satisfaction and safety in the automotive industry. Despite advancements in understanding various noise sources and mitigation strategies, vehicle noise continues to contribute to driver and passenger discomfort, impacting stress levels, fatigue, and overall quality of life. Recent research has made significant strides in classifying in-vehicle noise, yet the complexity of obtaining comprehensive and diverse datasets remains a major hurdle, given the variability and transient nature of these noises. To overcome these challenges, our research introduces an innovative approach using Few-shot Learning (FSL). We propose a unique FSL model that integrates a Triplet-trained Prototypical Network for the classification of in-vehicle noises. This model is particularly adept at learning robust feature representations from limited data. The application of triplet sampling and loss significantly enhances the model's ability to distinguish between various types of in-vehicle noises. Our methodology was rigorously tested using a specially curated dataset of in-vehicle noises, reflecting real-world diversity. The experimental results, obtained through 10-fold cross-validation, demonstrate an exceptional average accuracy of 96.81% on a 9-way 1-shot task. This level of accuracy, achieved with a limited amount of training data, not only attests to the effectiveness of our model but also marks a significant advancement in the field of acoustic classification. Our study's findings highlight the potential of FSL in addressing complex challenges in the automotive industry, paving the way for more effective noise reduction strategies and improved vehicle design.

INDEX TERMS Acoustic classification, representation learning, few-shot learning (FSL), in-vehicle noise, prototypical network, triplet loss.

I. INTRODUCTION

Vehicle noise, a critical concern in the automotive industry, significantly impacts customer satisfaction and poses potential safety hazards [1]. The implications of different pavement types on in-vehicle noise and their potential effects on human health have been highlighted by Li et al. [2]. Extensive research in in-vehicle noise has unearthed numerous sources and fostered innovative methods for mitigation. For instance, Alt et al. [3] introduced this field with an interior noise simulation, which was instrumental in predicting and refining the overall interior noise landscape, particularly focusing on the nuances of powertrain noise. Sang-Hyun [4] performed analytical noise and vibration simulations for design

The associate editor coordinating the review of this manuscript and approving it for publication was Jad Nasreddine.

implementation. Complementing this, Cornish [5] offered a groundbreaking methodology aimed at optimizing and stabilizing the quality of interior noise, emphasizing the crucial aspect of aligning with manufacturing realities.

These pioneering studies underscore the paramount importance of a nuanced understanding and strategic addressing of various in-vehicle noise sources to enhance the driving experience's comfort and enjoyment. Typical noise sources in vehicles include, but are not limited to, road noise, powertrain vibrations, braking systems, power electronics, and aeroacoustics. Each of these sources contributes uniquely to the overall noise level within the vehicle, affecting the quality impression of the vehicle. Despite significant advances in research aimed at mitigating these noises, in-vehicle noise continues to be a challenging issue, emphasizing the need for continued efforts in this area. The persistent presence of such noises underscores the urgency and significance of addressing this complex issue, given its extensive implications for both the user experience and the longevity of the vehicle.

In recent years, there has been a notable surge in studies focusing on the classification of in-vehicle noise. Huang et al. [6] implemented deep neural networks (DNNs) to predict shock absorber squeak noise. Data augmentation by empirical mode decomposition on neural networks was used to classify in-vehicle noises by Nam et al. [7]. Meanwhile, Bu and Cho [8] introduced an advanced deep beamforming network, surpassing existing deep learning algorithms with an impressive accuracy of 0.9270. In addition to this, Won [9] developed adaptive audio classification systems, enhancing accuracy across varied driving environments. These advancements are particularly commendable, given the formidable challenge traditionally posed by classifying diverse in-vehicle noises.

Traditionally, classifying these diverse in-vehicle noises has been a formidable challenge. The primary obstacle lies in the requirement for extensive, well-labeled datasets that cover the wide spectrum of potential noise sources, vehicle types, and environmental conditions. Compiling such a dataset is not only labor-intensive but also incurs significant costs. Furthermore, the conventional deep learning models used for acoustic classification are often data-hungry, necessitating large quantities of labeled samples to train effectively. This requirement becomes a major hindrance, especially when dealing with the vast variability of in-vehicle noises. The unique characteristics of the indoor environment cause sound to reflect, reverberate, and refract, and the distinctive features of the gearbox components inside the vehicle cause noises to occur abruptly and temporarily, resulting in very short data collection times [10]. Every year, various new vehicles are released, and new noise sources appear inside as time passes, making data collection very difficult.

To circumvent these challenges, our research pivots toward the cutting-edge field of Few-shot Learning (FSL). FSL, a paradigm in machine learning, emphasizes the development of models that can learn and make accurate predictions from a limited dataset [11], [12], [13], [14], [15], [16], [17], [18]. This methodology is exceptionally suited to scenarios like in-vehicle noise classification, where obtaining large and diverse datasets is impractical. By leveraging FSL, we aimed to create a model that can effectively generalize from a small sample of data, thus overcoming the traditional data constraints.

This study introduces a novel FSL model for in-vehicle noise classification, which incorporates a Triplet-trained Prototypical Network shown in Figure 1. This model distinguishes itself by adeptly learning robust feature representations from limited data. Employing STFT-based Mel-spectrogram embeddings, it effectively extracts key features from raw noise data, retaining critical information while simplifying the data's complexity [19], [20]. The application of triplet loss is a significant facet of our approach, enhancing the model's ability to discern between different in-vehicle noises. By minimizing the distance between similar noises and maximizing it between dissimilar ones, the model learns to identify subtle acoustic differences with remarkable efficiency [21], [22]. The incorporation of a prototypical network furthers our methodology's efficacy. This network excels in learning a metric space where classification is performed based on the proximity of embedded features to the prototype representation of each label [14].

This study leverages a specially curated dataset of in-vehicle noises, recorded under a variety of vehicles to mirror real-world diversity. By conducting rigorous evaluations, we seek to not only validate the efficacy of our proposed model but also contribute significant advancements to the field of acoustic classification. This research is composed to offer insights into the potential of FSL methodologies in achieving high accuracy in noise classification tasks, highlighting its advantages over traditional deep learning models and existing FSL approaches, particularly in scenarios characterized by data scarcity.

II. RELATED WORKS

FSL is a compelling paradigm addressing the data scarcity problem prevalent in various machine learning domains. The Model-Agnostic Meta-Learning (MAML) framework has been a pioneering force in this area, renowned for its versatility and model-agnosticism. MAML's approach to meta-learning, wherein models are trained to swiftly adapt to new tasks with minimal data, has significantly advanced FSL across multiple disciplines, including image classification and reinforcement learning [11]. It is particularly noted for its ability to fine-tune models rapidly, as evidenced by its state-of-the-art performance in image classification benchmarks.

Graph Neural Networks (GNNs) have further enriched the FSL landscape, offering a novel perspective through the lens of graphical model inference. By synthesizing traditional message-passing algorithms with neural network architectures, GNNs have birthed a framework that not only excels in numerical performance but also adapts seamlessly

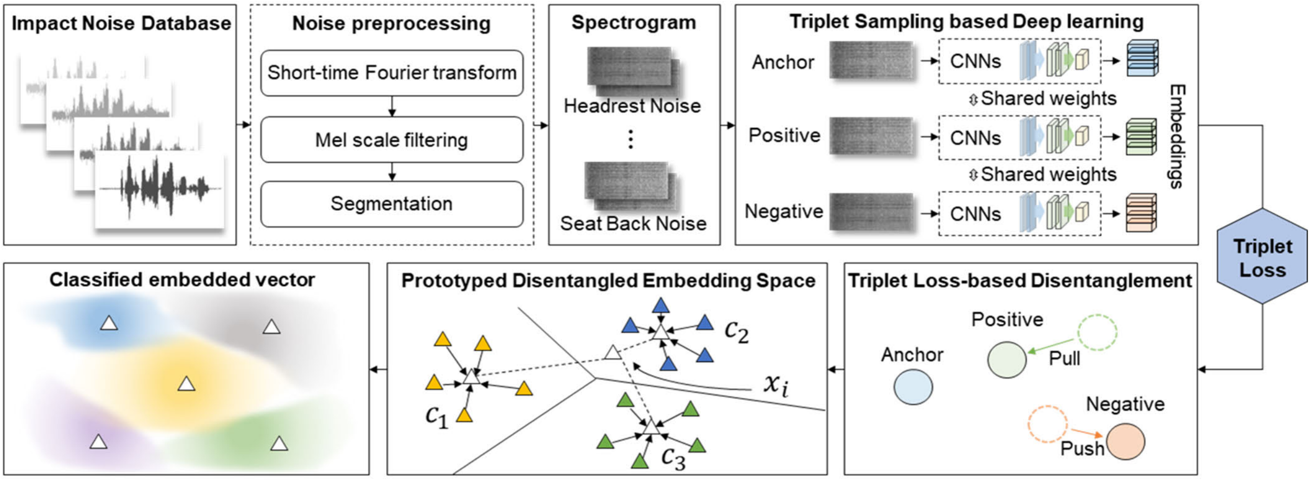


FIGURE 1. Overview of Triplet-trained prototypical network based FSL: collection of noise data, transformation into Mel-spectrogram embeddings, triplet loss-based disentanglement process, subsequent classification in a learned feature space.

TABLE 1. FSL methods for classification tasks under limited data conditions.

Method	Domain (dataset)	Task	Few-shot Accuracy
Model-Agnostic Meta-Learning [11]	Image (mini-Imagenet, Omniglot)	5-way 1-shot, 5-way 5-shot	63.1% 98.8%
Graph Neural Networks [13]	Image (mini-Imagenet, tiered-Imagenet, CIFAR-FS, CUB-200-2011)	5-way 1-shot, 5-way 5-shot	69.3% 84.2%
Memory-Augmented Nueral Networks [12]	Image (Omniglot, mini-Imagenet)	5-way 1-shot	95.2%
Prototypical loss [14]	Image (Omniglot, miniImageNet)	5-way 1-shot, 5-way 5-shot	96.0% 98.8%

to FSL variations such as semi-supervised and active learning [12]. Their relational task proficiency has been pivotal in expanding the boundaries of FSL applications. Frog-GNN, introduced by Xu and Xiang [23] combines a pre-trained language model and GNN, outperforming existing few-shot approaches in both few-shot text classification and relation classification. This further demonstrates the versatility and potential of GNNs in the domain of FSL.

In addressing the challenges of one-shot learning, Memory-Augmented Neural Networks (MANNs) have emerged a promising solution. Equipped with enhanced memory capabilities, these architectures exhibit remarkable efficiency in data assimilation and retrieval, thereby mitigating the limitations of conventional gradient-based models [13]. This leap in rapid data encoding and utilization marks a significant departure from the iterative training that plagues traditional networks. Generalized model-agnostic meta-learning (GMAML), presented by Lin et al. [24], involves constructing a channel interaction feature encoder using multi-kernel efficient channel attention, and has improved the overall generalization performance in various few-shot cross-domain scenarios. This further illustrates the potential of MANNs in the domain of FSL.

Prototypical Networks have played a pivotal role in FSL particularly in few-shot classification tasks, with their simple yet effective approach that relies on learning a metric space [14]. The study in [25] innovates further by integrating causal intervention with these networks, enhancing their performance in relation classification tasks. In [26] a specialized adaptation of Prototypical Networks significantly boosts few-shot handwritten character recognition for Urdu, a language with limited resources. Moreover, [27] illustrates the application of these networks in the culinary domain, integrating attention mechanisms for more robust food image recognition. He et al. [28] proposed virtual prompt pre-training method for prototype-based few-shot relation extraction, delivering a novel learning paradigm to model entities and relations via the probability distribution and Euclidean distance of the predictions of query instances and prototypes. Recently, a novel approach, SSL-ProtoNet, has been proposed by Lim et al. [29], which leverages self-supervised learning, Prototypical Networks, and knowledge distillation to enhance sample discrimination in few-shot learning tasks. Also, Wang et al. [30] introduces an innovative application of Prototypical Networks in FSL, where it leverages the prototypical concept to calibrate and enhance the discriminability of new class prototypes effectively.

The culmination of these approaches has paved the way for our work, where we aim to leverage the strengths of FSL in the nuanced domain of in-vehicle noise classification. By integrating a Disentangled Prototypical Network trained via Triplet Loss, our method not only aims to excel in classification accuracy but also maintain scalability and efficiency within the challenging automotive industry landscape.

This research echoes the advancements made by predecessors while extending their principles to the auditory realm, marking a novel contribution to both the fields of acoustic signal processing and FSL.

Algorithm 1 Triplet Training for Disentanglement of Embedding Space

This algorithm is designed to train a neural network to learn to disentangle the embedding space optimized by triplet loss. It generates triplets of anchor, positive, and negative samples and optimizes the network to bring the anchor closer to the positive sample while pushing it away from the negative sample in the embedding space.

Input:
X: Preprocessed impact noise data
Y: Corresponding labels

Output:
Z: Trained model that disentangles the embedding space optimized by triplet loss.

1: Generate triplets using X and Y, consisting of an anchor, a positive sample (same label as anchor), and a negative sample (different label from anchor).
2: Compute triplet loss for each triplet, considering the distance between anchor-positive and anchor-negative pairs in the feature space.
3: Create a convolutional neural network for feature extraction, taking X as input and returning a feature vector.
4. Train and optimize the network to minimize triplet loss, resulting in a model that generates distinct embeddings for different labels.
5: Return the model Z.

III. METHOD

A. IMPACT NOISE PREPROCESSING

In our approach to in-vehicle noise classification, the initial step involves a meticulous noise preprocessing routine. This process commences with the Short-time Fourier Transform (STFT) to convert time-domain impact noise data into the frequency domain, with a window length of 2048 frames (0.08 seconds) and overlapping samples of 512 steps. Which facilitated the extraction of spatiotemporal features.

Subsequent application of a Mel Filter Bank (Mel-fb) emphasizes perceptually relevant frequencies, aligning the data representation. Finally, sliding window segmentation was applied with a size of 16 to structure the noise data into consistent, analyzable segments, enhancing the model's capacity to learn from these transient noise events. The culmination of this preprocessing is a detailed noise spectrogram, serving as the input for the FSL model. This preparatory stage is crucial, transforming raw audio signals into a refined form suitable for advanced analysis, which is illustrated in Figure 2, visually showcasing the result of the preprocessing of the impact noise data.

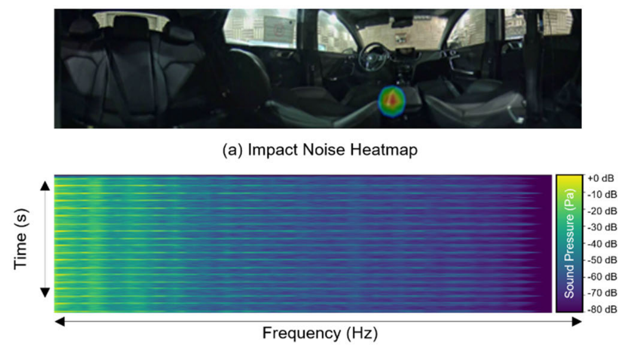
Algorithm 2 Prototypical Network for Disentangled Embedding Space FSL

This algorithm is designed for performing FSL on a disentangled embedding space using prototypical networks. It creates prototypes for each label using the support set and computes the distance between the query samples and the prototypes in the embedding space. It predicts the labels of the query samples based on the nearest prototype.

Input:
x_support: Support set of impact noise data
y_support: Corresponding labels of the support set
x_new: Query set for prediction of impact noise data
T_model: Triplet-Trained embedding model

Output:
Y: Prediction of corresponding labels of the query set.

1: Embed x_support set using T_model and obtain the feature vectors.
2: For each label, compute the mean of the feature vectors of the x_support samples belonging to that label. These are the prototypes.
3: Embed the x_new set using the T_model and obtain the feature vectors.
4: For each x_new sample, calculate the Euclidean distance to each prototype.
5: Assign the label of the nearest prototype to the x_new sample.
6: Return the predicted label Y.



(a) Impact Noise Heatmap

(b) Preprocessed Noise Spectrogram

FIGURE 2. Visual representation of noise preprocessing. (a) displays the impact noise heatmap captured within the vehicle interior. (b) shows the resulting preprocessed noise spectrogram.

B. TRIPLET LOSS-BASED DISENTANGLEMENT

Before the FSL process, the embedding space of in-vehicle noise features is disentangled via a deep learning architecture employing triplet sampling. The method harnesses a triplet sampling mechanism, where each triplet comprises an anchor, a positive sample with the same label as the anchor, and a negative sample with a different label. The feature embedding function learned by the network $f(\cdot)$ processes these instances, projecting them into an embedding space where the triplet loss function takes effect. The triplet loss function is defined as equation (1).

$$T_{loss} = \sum_{i} \left[\| f(a_i) - f(p_i) \|^2 - \| f(a_i) - f(n_i) \|^2 + \alpha \right]_+ \quad (1)$$

where Σ_i is a summation of all the triplets in the batch. It quantifies the distance between an anchor instance $f(a_i)$ and a positive instance $f(p_i)$, which shares the same label, in contrast to a negative instance $f(n_i)$, which has a different label. The parameter α is a margin that enforces a minimum gap between positive and negative pairs. The L2 norm squared $\|\cdot\|^2$ calculates the squared Euclidean distance between two embeddings. To ensure the loss remains non-negative, $[\cdot]_+$ is a hinge function, which is equivalent to $\max(\cdot, 0)$.

The anchor aims to minimize the distance to the positive while maximizing the distance from the negative, thus disentangling the embedding space. This loss facilitates a learning dynamic that not only focuses on correct classification but also on the relative positioning of similar and dissimilar examples, thus enhancing the model's discriminative ability.

This strategic approach underscores the balance between precision in classification and robustness against dataset variability, crucial for real-world application in acoustic signal processing.

C. NOISE PROTOTYPING FOR FSL

In the domain of FSL, the Prototypical Network framework forms the backbone of our methodology, which is adept at learning a metric space where classification is performed based on the proximity to prototype representations of each label. The process begins with the generation of prototypes for each label through the averaging of embedded features from support queries shown in equation (2). This yields a centroid that embodies the essence of each label within the embedding space. Subsequently, classification of a new query involves computing its distance to these prototypes, utilizing a Euclidean distance function, equation (3). The query is then assigned the label of the nearest prototype, thus effectuating the classification, equation (4).

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\phi(x_i) \quad (2)$$

$$d(x_{new}, p_i) = \sqrt{\sum (x_{new} - p_i)^2} \quad (3)$$

$$\hat{y} = \arg\min_i d(x_{new}, p_i) \quad (4)$$

Here, x_i, y_i is a set of support queries with its corresponding label, and c_k is the prototype of the k^{th} label. f_ϕ is the embedding function which is made from the deep learning process of Part B, and d is the Euclidian distance function to calculate the distance between a new query sample x_{new} and the prototype p_i corresponding to its label. $\hat{y}$ represents the predicted label for x_{new}.

This strategy is particularly potent in scenarios where data is scarce, and yet, the embedding space is richly structured by prior triplet loss disentanglement, ensuring a well-organized and discriminative feature landscape for accurate classification.

IV. EXPERIMENTAL RESULTS

A. IN-VEHICLE NOISE DATASET IN REAL-WORLD

To preserve the frequency characteristics of the impact noise that is propagated along the vehicle steering system, the sampling process is conducted in the faulty vehicle. One of the globally renowned motor groups collected 9 noise types from 22 different vehicles including sedans and SUVs under the same sampling condition shown in Table 2. A 25,600 Hz sampling rate is chosen, a specification for common smartphone microphones on the market, to ensure practicality. Each noise sample is 2 seconds long and contains 10 to 12 impact signals.

TABLE 2. Sample length and number of in-vehicle noise data.

Index	Noise Type	Sample Length	Sample Count
0	Etc.	2 sec	1200
1	Headrest Noise		3325
2	Door Trim Noise		3552
3	Driver Seat Airbag Noise		1344
4	Center Console Noise 1		1152
5	Seat Back Noise		960
6	Center Console Noise 2		336
7	Left Side Mirror Noise		48
8	Right Side Mirror Noise		1920

Since 10-fold cross validation was used to prove robustness, 90% of the total data for each fold was used for the triplet train, and 10% was used as test data, that is, for prototyping and prediction.

B. CLASSIFICATION PERFORMANCE WITH LIMITED DATA

The experimental results shown in Table 3 demonstrate the superior performance of the proposed model in in-vehicle noise classification under limited data conditions. Hyper-parameters of all models were not optimized in these experimental results.

TABLE 3. 9-Way FSL average accuracy (%).

Few-shot Method	9-Way 10-Shot	9-Way 5-Shot	9-Way 1-Shot
CNN	80.85±10.33	79.47±8.31	75.41±8.14
MANN	86.75±1.47	83.29±1.61	63.72±3.49
GNN	87.47±0.90	87.82±2.15	81.80±3.97
Prototypical Networks	94.48±1.65	94.02±1.33	92.14±3.49
Proposed	99.57±1.27	99.51±0.82	96.81±7.41

In a 9-Way 10-Shot and 9-Way 5-Shot setup, our model achieved an exceptional 99.57% and 99.51% accuracy, surpassing traditional methods such as CNNs, MANNs, GNNs, and Prototypical networks. The robustness of the model is further evidenced in the challenging 9-Way 1-Shot scenario, where it maintained a high accuracy of 96.81%.

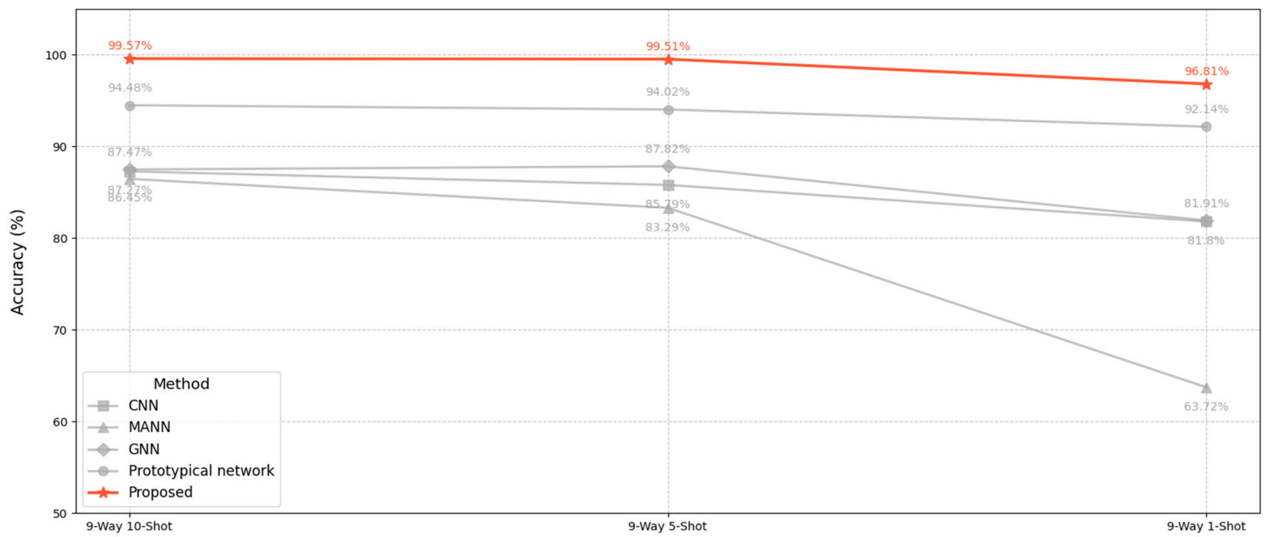


FIGURE 3. **Accuracy degradation comparison according to the data limits. This graph illustrates the classification accuracy of the proposed model against traditional methods across different few-shot scenarios in a data-decreasing order, from 9-Way 10-Shot to 9-Way 1-Shot.**

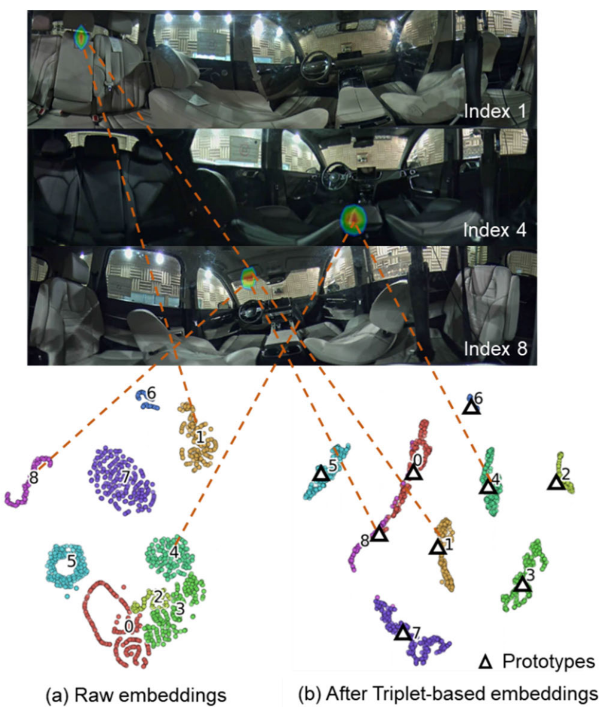


FIGURE 4. **Visualization of embedding space disentanglement. (a) exhibits the raw embeddings from in-vehicle noise data, depicting a high degree of overlap among different noise types. (b) illustrates the clarity achieved in the embedding space after applying the triplet-based embeddings.**

The degradation of data limit is well shown in Figure 3. These results confirm the efficacy of the FSL approach in accurately classifying noise with minimal data, showcasing the model's potential for practical applications where data is sparse.

C. DISENTANGLEMENT OF TRIPLET-TRAINED MODEL

Our triplet-based disentanglement method significantly improves how we classify in-vehicle noise with FSL. Figure 4 clearly shows how effective this method is. Before using our method (shown in part (a) of Figure 4), the noise features are mixed together, making it hard to tell different noises apart. This entanglement makes accurate classification difficult, especially when noises are subtle or similar.

After applying our triplet-trained embeddings, as shown in part (b) of Figure 4, there's a remarkable change. The noise features are now separated clearly, making it much easier to identify and classify different types of noises. This clear separation is crucial for our model to classify noises accurately, especially when it has only a few examples to learn from.

What makes our approach stand out is its ability to handle the complexity and variety of noises inside vehicles better than previous methods. Traditional methods often struggle with similar sounding noises, but our approach can distinguish these subtle differences reliably.

V. CONCLUSION

The study culminated in a robust FSL model adept at classifying in-vehicle noise with minimal data. Notably, it outperformed existing methodologies in a 9-Way 1-Shot scenario with an accuracy of 96.8%, showcasing the model's proficiency with limited datasets. The method's success is attributed to the novel integration of a triplet loss-based disentanglement in conjunction with a prototypical network, leading to substantial accuracy improvements.

Reflecting on the work's implications, it becomes clear that this research marks a significant leap in few-shot in-vehicle noise classification. It confronts the prevalent data acquisition challenges and efficiently handles the diversity of noises within vehicles. The approach's effectiveness is underscored by a comparison with several established models, where it consistently delivers superior performance.

However, the research acknowledges certain limitations, including the need for further validation of noisy datasets. Also, the hyperparameter optimization was not conducted for any of the methods evaluated, including our proposed model as well as the traditional methods (CNN, MANN, GNN, Prototypical Networks). This decision was made to ensure a consistent and uniform comparison framework across all methods.

We recognize, however, that this approach may impact the representativeness of the reported accuracy figures, as hyperparameter tuning has the potential to significantly enhance the performance of each method. Also, uniform noise sampling conditions and the absence of varied environmental acoustic disturbances in the training set could include potential biases.

Moving forward, the future work will focus on enhancing the model's robustness by diversifying the noise conditions, including urban sounds, and varying vehicle operational states, implementing a continual learning framework to adapt to new noise data seamlessly, explore hyperparameter optimization, and extending the model's applicability to real-world diagnostic scenarios. This progression aims not only to validate the model's efficacy further but also to translate the academic findings into practical solutions for vehicle condition improvement.

REFERENCES

[1] M. S. Qatu, "Recent research on vehicle noise and vibration," *Int. J. Vehicle Noise Vibrat.*, vol. 8, no. 4, p. 289, 2012.

[2] Q. Li, F. Qiao, and L. Yu, "Impacts of pavement types on in-vehicle noise and human health," *J. Air Waste Manage. Assoc.*, vol. 66, no. 1, pp. 87–96, Jan. 2016.

[3] N. W. Alt, N. Wiehagen, and M. W. Schlitzer, "Interior noise simulation for improved vehicle sound," SAE, Warrendale, PA, USA, Tech. Paper 01-1539, 2001.

[4] J. Sang-Hyun, "The application of the simulation techniques to reduce the noise and vibration in vehicle development," in *Proc. FISITA World Automot. Congr.*, Jun. 2000, pp. 12–15.

[5] R. Cornish, "A novel approach to optimizing and stabilizing interior noise quality in vehicles," *Proc. Inst. Mech. Engineers, D, J. Automobile Eng.*, vol. 214, no. 7, pp. 685–692, Jul. 2000.

[6] H. B. Huang, X. R. Huang, J. H. Wu, M. L. Yang, and W. P. Ding, "Novel method for identifying and diagnosing electric vehicle shock absorber squeak noise based on a DNN," *Mech. Syst. Signal Process.*, vol. 124, pp. 439–458, Jun. 2019.

[7] G.-H. Nam, S.-J. Bu, N.-M. Park, J.-Y. Seo, H.-C. Jo, and W.-T. Jeong, "Data augmentation using empirical mode decomposition on neural networks to classify impact noise in vehicle," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 731–735.

[8] S.-J. Bu and S.-B. Cho, "Classifying in-vehicle noise from multi-channel sound spectrum by deep beamforming networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 3545–3552.

[9] M. Won, H. Alsaadan, and Y. Eun, "Adaptive multi-class audio classification in noisy in-vehicle environment," 2017, *arXiv:1703.07065*.

[10] J.-T. Kim, S. M. Lee, T. Lee, and W.-G. Kim, "Study on mechanism of impact noise on steering gear while turning steering wheel in opposite directions," in *Proc. Inter-Noise Noise-Con Congr. Conf.*, 2016, vol. 253, no. 7. Washington, DC, USA: Institute of Noise Control Engineering, pp. 1095–1105.

[11] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 1126–1135.

[12] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–12.

[13] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 1842–1850.

[14] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.

[15] H. Lin, G. Han, J. Ma, S. Huang, X. Lin, and S.-F. Chang, "Supervised masked knowledge distillation for few-shot transformers," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 19649–19659.

[16] S. J. Reddi, "Efficient training of language models using few-shot learning," in *Proc. Int. Conf. Mach. Learn.*, 2023, pp. 14553–14568.

[17] H. Cheng, S. Yang, J. T. Zhou, L. Guo, and B. Wen, "Frequency guidance matters in few-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 11814–11824.

[18] P. Mazumder, P. Singh, and P. Rai, "Few-shot lifelong learning," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 3, pp. 2337–2345.

[19] A. Graves, A.-R. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6645–6649.

[20] A. Graves and N. Jaitly, "Towards end-to-end speech recognition with recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1764–1772.

[21] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *Proc. Int. Workshop Similarity-Based Pattern Recognit. (SIMBAD)*, Copenhagen, Denmark. Berlin, Germany: Springer, 2015, pp. 84–92.

[22] S.-J. Bu, N. Park, G.-H. Nam, J.-Y. Seo, and S.-B. Cho, "A Monte Carlo search-based triplet sampling method for learning disentangled representation of impulsive noise on steering gear," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 3057–3061.

[23] S. Xu and Y. Xiang, "Frog-GNN: Multi-perspective aggregation based graph neural network for few-shot text classification," *Expert Syst. Appl.*, vol. 176, Aug. 2021, Art. no. 114795.

[24] J. Lin, H. Shao, X. Zhou, B. Cai, and B. Liu, "Generalized MAML for few-shot cross-domain fault diagnosis of bearing driven by heterogeneous signals," *Exp. Syst. Appl.*, vol. 230, Nov. 2023, Art. no. 120696.

[25] Z. Li, F. Ouyang, C. Zhou, Y. He, and L. Shen, "Few-shot relation classification research based on prototypical network and causal intervention," *IEEE Access*, vol. 10, pp. 36995–37002, 2022.

[26] R. Sahay and M. Coustaty, "An enhanced prototypical network architecture for few-shot handwritten Urdu character recognition," *IEEE Access*, vol. 11, pp. 33682–33696, 2023.

[27] G. Song, Z. Tao, X. Huang, G. Cao, W. Liu, and L. Yang, "Hybrid attention-based prototypical network for unfamiliar restaurant food image few-shot recognition," *IEEE Access*, vol. 8, pp. 14893–14900, 2020.

[28] K. He, Y. Huang, R. Mao, T. Gong, C. Li, and E. Cambria, "Virtual prompt pre-training for prototype-based few-shot relation extraction," *Exp. Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 118927.

[29] J. Y. Lim, K. M. Lim, C. P. Lee, and Y. X. Tan, "SSL-ProtoNet: Self-supervised learning prototypical networks for few-shot learning," *Exp. Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122173.

[30] Q. W. Wang, D. W. Zhou, Y. K. Zhang, D. C. Zhan, and H. J. Ye, "Few-shot class-incremental learning via training-free prototype calibration," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 1–9.

ROBIN INHO KEE received the B.S. degree in mechanical engineering from Yonsei University, Seoul, South Korea, in 2022. He is currently pursuing the M.S. degree in mechanical engineering with the University of Michigan, Ann Arbor, MI, USA. From 2022 to 2023, he was a Research Intern with the Center for Intelligent and Interactive Robotics, Korea Institute of Science and Technology, Seoul. His research interest includes the development of deep learning models and control methods for wearable robots.

DAHYUN NAM is currently pursuing the B.S. degree in mechanical engineering with Yonsei University, Seoul, South Korea. She is currently a Research Intern with the BioRobotics Laboratory, Seoul National University, Seoul. Her research interest includes the development of deep learning models and designs for mobile robots.

SEOK-JUN BUU (Member, IEEE) received the Ph.D. degree in computer science from Yonsei University, South Korea. Since 2023, he has been an Assistant Professor with the Department of Computer Science, Gyeongsang National University. His primary research interest includes the application of deep learning to address a wide array of industrial challenges through empirical methods. He is focused on enhancing deep learning applications by incorporating domain knowledge and real-world constraints, as well as optimizing architecture and hyperparameters through traditional AI techniques, such as genetic algorithms.

SUNG-BAE CHO (Senior Member, IEEE) received the B.S. degree in computer science from Yonsei University, Seoul, South Korea, and the M.S. and Ph.D. degrees in computer science from KAIST, Daejeon, South Korea. He was an Invited Researcher of Human Information Processing Research Laboratories, ATR Institute, Kyoto, Japan, from 1993 to 1995, and a Visiting Scholar with the University of New South Wales, Canberra, Australia, in 1998. He was a Visiting Professor with The University of British Columbia, Vancouver, Canada, from 2005 to 2006, and the King Mongkut's University of Technology at Thonburi, Bangkok, Thailand, in 2013. Since 1995, he has been a Professor with the Department of Computer Science, Yonsei University, a Underwood Distinguished Professor, since 2021, and a Yonsei Lee Youn-Jae Fellow, since 2023. His research interests include neural networks, pattern recognition, intelligent man–machine interfaces, evolutionary computation, and artificial life. He was a recipient of the Richard E. Merwin Prize from the IEEE Computer Society, in 1993. He was also a recipient of the Service Merit Medal from Korean Government, in 2022. He received several Distinguished Investigator Awards from Korea Information Science Society, in 2005, and the Gaheon Sindoricoh, in 2017.