World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:16, No:8, 2022

A Multi-Feature Deep Learning Algorithm for Urban Traffic Classification with Limited Labeled Data

Rohan Putatunda, Aryya Gangopadhyay

***Abstract*—**Acoustic sensors, if embedded in smart street lights, can help in capturing the activities (car honking, sirens, events, traffic, etc.) in cities. Needless to say, the acoustic data from such scenarios are complex due to multiple audio streams originating from different events, and when decomposed to independent signals, the amount of retrieved data volume is small in quantity which is inadequate to train deep neural networks. So, in this paper, we address the two challenges: a) separating the mixed signals, and b) developing an efficient acoustic classifier under data paucity. So, to address these challenges, we propose an architecture with supervised deep learning, where the initial captured mixed acoustics data are analyzed with Fast Fourier Transformation (FFT), followed by filtering the noise from the signal, and then decomposed to independent signals by fast independent component analysis (Fast ICA). To address the challenge of data paucity, we propose a multi feature-based deep neural network with high performance that is reflected in our experiments when compared to the conventional convolutional neural network (CNN) and multi-layer perceptron (MLP).

***Keywords*—**FFT, ICA, vehicle classification, multi-feature DNN, CNN, MLP.

I. Introduction

UNDERSTANDING traffic network flow volume at various locations, particularly at road intersections, can be very useful in city planning. For example, such information can be used to design or modify existing traffic lights, broaden or improve the quality of certain roads, as well as build additional roadways to improve traffic circulation. Since streetlights are ubiquitous in most cities acoustic sensors installed in the streetlights are ideal for capturing the surrounding audio signals. These signals can subsequently be analyzed to identify the volume and nature of the traffic and their diurnal patterns, for example, rush hour versus non-rush hour traffic. However, there are several challenges in the classification of the specific types of acoustic signals. First, most of the acoustic signals are a mixture of various traffic and other activities in the surrounding area. The second big challenge, as in many other classification tasks, is the absence of enough labeled datasets, which makes it challenging to train machine learning models. In this paper, we present a methodology with a multi-feature deep learning model to address the data paucity challenge to build an efficient acoustic classifier.

We capture mixed acoustic signals from different road intersections in the city area, detect the number of individuals' signals/frequencies present in the mixed acoustics with the help of the FFT and then implement the ICA to extract the individual signals from the mixed acoustics.

Next, we train a deep learning model to classify each extracted signal into one of three vehicle types: light (i.e., car), medium (i.e., bus), or heavy (i.e., truck). Audio signals have several advantages as compared to video and image signals in terms of storage, data compression, and localization. Perhaps the biggest advantage of audio signals over the others mentioned above is that of data privacy. Video or images may accidentally capture the facial images of the drivers which can pose a threat to their privacy. As opposed to that audio signals do not contain any privacy-sensitive information. We then evaluate the performance of our acoustic classifier with the validation accuracies, losses, and three other evaluation parameters of the classification report which are F-1 score, precision, and recall. One of the big challenges in deep learning is the lack of enough labeled data. In our experimental phase, we observe that standard data augmentation techniques such as adding noise, shifting, and stretching the signals have not resulted in improvements in model performance. Our paper addresses the problem of developing an efficient classifier under a data paucity scenario. The methodologies that are adopted for the same are a) multi-feature learning and b) knowledge transfer from cross-domain adaptation to improve the performance of the classifier.

The rest of the paper is organized as follows: in Section II we discuss the related research; in Section III we describe proposed methodology; in Section IV we describe experimental results, and in Section V we provide conclusions and ideas for future work.

II. Related Work

Acoustic signal processing for vehicle detection and classification has not been used extensively in city planning applications, although there is great potential in finding traffic patterns in such applications. Previously tri-stimulus response features used in the analysis of music theory are being used for the characterization of the car engine sound [1]. In another work, quadratic discriminant analysis was used for various types of vehicle classification. This approach was based on considering energy vectors [2]. Another recent work focused on the detection of vehicles by the modification of the microphone signals spectral processing algorithm along with noise filtering. In this work, they detect moving road vehicles by using the records of acoustics signals [3]. Types of vehicle detection using emission harmonics have been developed for battlefield

Rohan Putatunda is with University of Maryland, Baltimore County, United States (e-mail: rohanp1@umbc.edu).

ground vehicles using the emitted acoustic signals captured by the wireless sensory network framework [4].

An intelligent transportation system has been developed for vehicle counting. The vehicle count algorithm used a dynamic time wrapping on the generated sound map [5]. A fusion-based learning approach was developed to classify the vehicles. Fusion refers to utilizing the audio and video signals, the classification is done on the data tuple, over here the traffic relevant features are extracted from both the audio and the image data. Information fusion at multiple levels (audio, image, and overall) shows consistently better performance [6]. As a part of the intelligent transportation system, an agent-based approach has been developed to manage the traffic network so that a non–intrusive grid of agent-based sensors is able to monitor the traffic parameters, in this case, it explores the various acoustic signature generated by different kind of vehicles which is then further utilized for estimating the traffic flows [7]. Other research in detecting, monitoring, and predicting urban traffic includes those using images [8], traffic data obtained from traffic control stations [9], and using the Internet of Things [10].

The related research on acoustic-based vehicle classification is based on the presence of a high volume of collected or captured acoustics data for training the classifier. The uniqueness of our work is based on training the efficient classifier from a limited amount of labeled data under a mixed acoustics scenario. The approach of multi-feature and cross-domain knowledge transfer function is integrated with the deep neural network to improve the performance of the model.

III. Methodology

In this section, we present the overall methodology for sensing acoustic data and classifying urban traffic which is reflected in Fig. 1 with our multi-feature-based deep learning model, in addition, each component of the overall methodology is discussed in detail in the subsection.

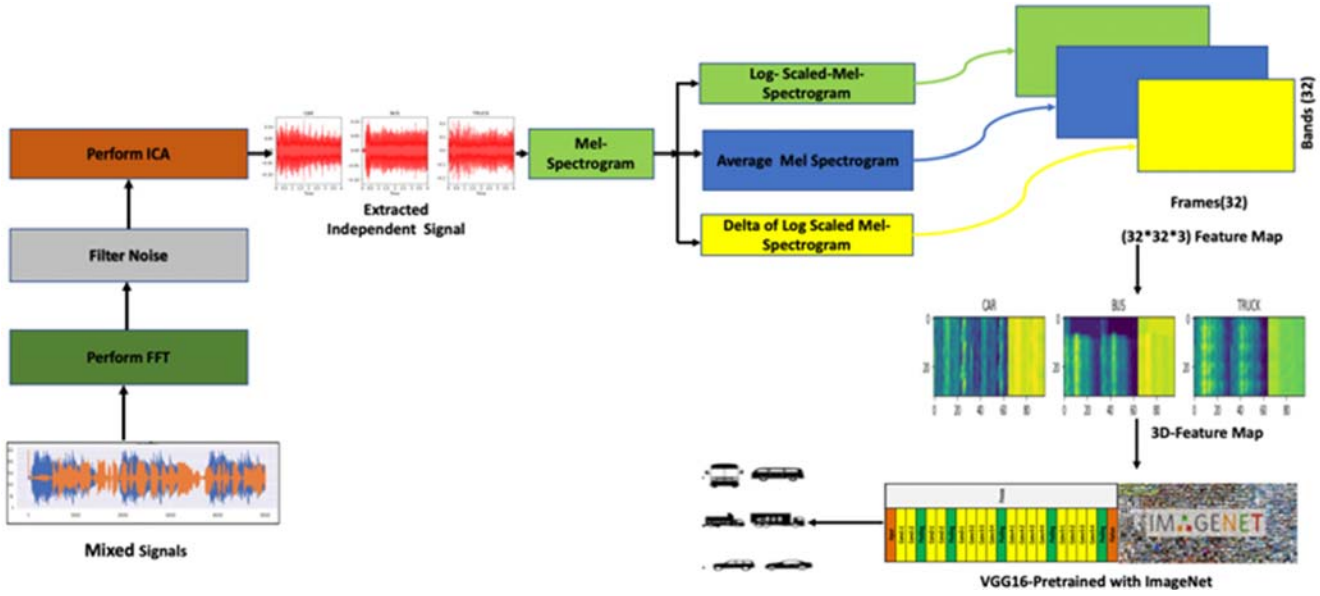

Fig. 1 Overall Methodology

A. Capturing Mixed Signals

Acoustic sensors are mostly placed on the streetlights to capture the signals to understand the urban traffic flow; to imitate a similar scenario we collected/recorded acoustics from the city audio traffic signals from various intersections at various times of the day. In order to capture the audio signals, the *Wave Editor* mobile application system was used. From the mobile phone, the *.wav* files of the acoustic signals were further analyzed. The platforms used for the experiments include the *audacity* tool and codes written with an object-oriented programming language (*python*) for the purpose of knowledge extraction from the raw audio signals. Fig. 2 depicts a typical traffic flow scenario with a combination of cars, trucks, and buses with random noise such as wind and rain. These kinds of random noise can sometimes drown the traffic acoustics but need to be filtered out from the mixture of audio signals before converting the mixed signals to independent component signals belonging to various vehicle classes.

B. Analyzing Mixed Signals with FFT

Before classifying the component audio signals from mixed signals, we needed to determine the number of component signals that are present in the mixed acoustic signal. To determine the number of individual signals we used FFT which clearly indicated the different frequencies that are present in them. This allowed us to convert a time-sequenced mixed signal into its frequency domain. We mostly observed the different patterns of frequencies in the captured mixed signals when analyzed with the FFT algorithm, which can be summarized with three scenarios below. In the scenario 1, most of the peaks in the amplitude domain with respect to the frequency domain were in the range between 0 to 1000 HZ, and the highest peak

in the signal corresponded to 1.5 dB, which is the acoustic signature generated by a truck engine followed by other vehicle engines. The lower profile amplitudes are the acoustics of pedestrians and steady wind flow. In scenario 2, upon implementing the FFT on the mixed signals that consisted of the acoustics of a bus and a truck with the emergency siren from a car suppressed the relevant acoustics with the peak of 2.5 dB in the frequency range of 0 to 500 Hz that need to be removed by filtering methodology to generate independent signals for better classification. Furthermore, most of the peak amplitudes of other vehicles along with the base amplitude of the steady flow of the wind are also observed in the frequency range of 0 to 1000 Hz, while scenario 3 highlights the captured acoustics consisting of the sound of cars, buses, and trucks, along with the heavy noise of a wind gust with the amplitude range of 2 dB in the frequency range of 500 Hz, and most of the peak amplitudes observed in the range of 0 to 1000 Hz. The main objective of using the FFT on top of the mixed signals is to identify the high-frequency range noise that can be filtered out from the acoustic signals which can increase the effectiveness of the ICA algorithm.

C. Filtering of Noise from Mixed Acoustics

In the filtering process, we studied the scenario where the truck is idling at a traffic signal along with a police car which is generating a siren that is irrelevant to our analysis. The peak amplitudes with reference to frequency signified the noise generated by the siren. To filter the random siren noise, we use the notch filter which will reduce the frequencies with respect to higher noises. We observed that after applying the notch filter in the 1300 Hz and 2200 Hz band the peaks of the acoustic signal were smoothed with a reduction in noise level. In our experiments, all the acoustic data cleaning is pre-processed by a notch filter to smooth any irrelevant noise such as a siren, a gust of wind, steady wind, and pedestrian sounds. After the filtering process, we performed an ICA in which we create the individual signals from the captured mixed acoustics which will be discussed in the next subsection.

D. Separating Mixed Signals with Fast ICA Algorithm

To convert the mixed acoustics signals into independent component signals we implemented the Fast ICA algorithm which works on two assumptions. The hidden component that we are trying to uncover must be statistically independent and second, it should be non-Gaussian. To initiate the Fast ICA algorithm, we first whiten the signal and then choose a random initial value for the de-mixing matrix(w) followed by choosing the new value of w with normalization. Finally, we took the dot product value between the de-mixing matrix and the signal value(x) to retrieve the independent signal. Fig. 3 shows the independent component signals that are separated from the mixed signals, and these very independent signals are used for extracting the relevant feature information for training an efficient classifier.

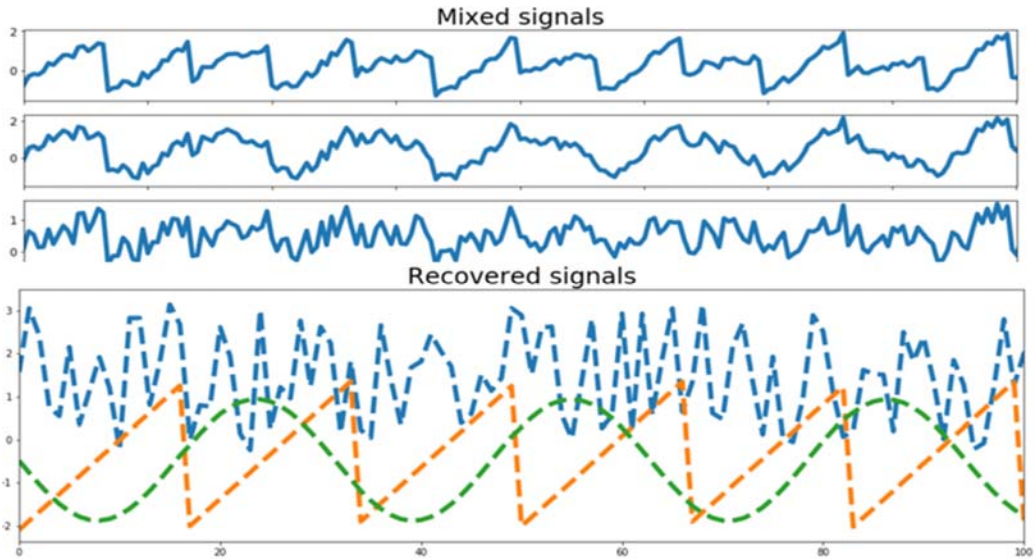

Fig. 2 Independent Component Signal by ICA Algorithm

E. Feature Extraction

In our work, we converted the extracted independent signals into the form of a mel-spectrogram, which is a better feature representation to train the deep learning models. The mel-spectrogram feature represents an acoustic time-frequency representation of a sound, the power spectral density function for the same is $P(f, t)$. It is sampled into a number of points around equally spaced times t_i and frequencies f_j (on a mel-frequency scale). In our case for the MLP and CNN, we only used the mel-spectrogram feature to train the model whereas in the case of the multi-featured deep learning model we converted the mel-spectrogram into three sub-features of log-scaled mel-spectrogram, an average of harmonic and percussive components mel-spectrogram, and delta of log scaled mel-spectrogram which forms a three-dimensional space feature representation which consists of higher information than normal mel-spectrogram.

IV. Experimental Results

A. Dataset

We captured nearly 112 mixed acoustic signals from various intersections of the city, across various times of the day. Each mixed signal has a length of between 10-11 seconds. After preprocessing with ICA of the mixed acoustic signals, we are being able to retrieve 50 independent acoustic signals of which 17 belong to the truck class followed by 12 belonging to the bus, and 11 belonging to the car.

B. Conventional Data Augmentation

Data augmentation is a necessary step for many deep neural network applications where there is a scarcity of labeled data. Data augmentation for images is performed by *flipping, cropping, rotation, and rolling* to convert the total volume of the dataset from 50 to 100 data points. In this research work, we observed that these small volumes of data are when augmented with conventional methodology result in poor classification, so to address the data paucity we introduce our methodology of multi-feature generation from a single audio signal with which we achieved higher performance for the classification model.

C. Conventional Model Architecture (CNN and MLP Model)

MLP model: The ANN model was a fully connected network with three hidden layers with 32 units in each of the layers. The activation functions for each hidden layer were RELU. In addition, we used dropout regularization with a dropout rate of 0.4. The activation function for the output layer was softmax with three classes: cars, buses, and trucks. The Adam optimization algorithm was used for a categorical cross-entropy loss function.

CNN model: The DNN model had two 2D convolution layers with 32 units and a kernel size of 3 with the "same" padding. Each convolutional layer was followed by a 2D max pooling with the same padding. The RELU activation was used in the hidden layers. The last layer before the last Conv2d was to flatten the output feature maps into one-dimensional vectors. After the flatten layer, there were two additional two fully connected hidden layers with RELU activation function. In the output layer, we used the softmax function for the three classes. The loss function was minimized using the Adam optimizer.

D. Multi-Feature with Cross-Domain Knowledge Transfer Deep Neural Networks

To develop an efficient vehicle detection algorithm based on raw acoustics data, we need good representations from our raw data. We have observed during our data collection phase that our captured acoustics have varying lengths. However, to build a robust classifier, our features need to be a consistent per sample. So, in order to have the fixed length, we extracted the audio sub-samples which have a standardized length, on which we extracted our mel-spectrogram features. In our work we extracted multi-feature representation from every single extracted sub-sample of acoustic data, in this paper we adopted three feature engineering techniques to build three feature representation maps, which output as a three-dimensional image feature map. The process is approached by converting audio sub-samples into mel-spectrogram and from this mel-spectrogram, we are able to formulate three multi-feature maps which are log scaled mel-spectrogram, average (harmonic and percussive) mel-spectrogram, and delta (derivative) of log scaled mel-spectrogram. In the next step for feature extraction, we utilize VGG-16 architecture which is pretrained on the weights of the ImageNet dataset to extract the bottleneck features which consist of the most abstracted information that plays a vital role in training an appropriate model with high inference results. Our deep neural network has four hidden layers. The first two hidden layers consist of 1024 units followed by another two hidden layers consisting of 512 units. The RELU activation function was used in all the four hidden layers. The dropout rate for first two hidden layers is .4 which is followed by .5 dropout for the other two hidden layers. The output layer has the softmax as the activation function with the loss as the categorical cross-entropy along with Adam optimizer to prevent the overfitting issues.

E. Model Performance Evaluation

The deep learning models are evaluated on five parameters which are reflected in Tables I and II.

Accuracy: Accuracy metrics in our experiment that is reflected in Table I signify the overall accuracy of the classifier to determine the fraction of the samples that are correctly classified by the classifier. In these metrics, we can observe that with the multi-feature representation integrated with cross-domain knowledge transfer deep learning model has a lot better performance in terms of accuracy when compared to the other models. It is also observed that conventional data augmentation has no impact on the performance improvement of the model.

TABLE I
Models Performance

ANN-50 Original Dataset										
Epoch	10	20	30	40	50	60	70	80	90	100
Loss	5.7	5.4	5.5	5.6	5.2	5.3	5.1	5.4	5.7	5.6
Accuracy	.4	.4	.41	.39	.43	.44	.44	.46	.42	.44
ANN-100 Original Dataset with Conventional Data Augmentation										
Epoch	10	20	30	40	50	60	70	80	90	100
Loss	5.8	5.7	5.9	6.1	5.8	5.7	5.6	5.7	5.8	5.9
Accuracy	.32	.39	.41	.43	.40	.37	.38	.41	.38	.37
CNN-50 Original Dataset										
Epoch	10	20	30	40	50	60	70	80	90	100
Loss	5.01	5.2	4.9	4.8	5.1	4.9	5.2	5.1	5.2	5.1
Accuracy	.45	.47	.46	.48	.52	.57	.56	.62	.58	.60
CNN-100 Original Dataset with Conventional Data Augmentation										
Epoch	10	20	30	40	50	60	70	80	90	100
Loss	5.35	5.4	5.1	5.4	5.7	5.6	5.5	5.3	5.6	5.6
Accuracy	.43	.48	.39	.45	.4	.43	.51	.54	.53	.56
Multi-Feature DNN-50 Original Dataset										
Epoch	10	20	30	40	50	60	70	80	90	100
Loss	.69	.58	.45	.31	.28	.23	.33	.32	.22	.30
Accuracy	.69	.75	.80	.88	.89	.90	.89	.90	.92	.90

Loss: In any deep learning model, the loss function is often referred to as the penalty for bad prediction. The loss function basically measures the performance of the model on a particular task. In our case, we achieved the most optimized loss function

with our proposed multi-feature deep learning models when compared to the other models.

Precision: Precision metrics in this context signify the fraction of predictions as the positive class that is originally classified as positive. In these metrics also made a similar kind of observation to overall accuracy metrics in which multi-feature-based DNN outperforms other models.

Recall: Recall is defined as the fraction in which all positive samples are correctly predicted as positive by the classifier in which we can observe a similar trend like the precision metrics.

F1-Score: The F1 score is defined as the harmonic mean of precision and recall. In our context, the multi-feature data augmentation-based DNN for scaling the data volume proved better results compared to conventional data augmentation.

TABLE II
CLASSIFICATION REPORT

ANN-50 Original Dataset		
Precision	Recall	F1-Score
.45	.43	.44
ANN-100 Original Dataset with Conventional Data Augmentation		
Precision	Recall	F1-Score
.37	.36	.365
CNN-50 Original Dataset		
Precision	Recall	F1-Score
CNN-100 Original Dataset with Conventional Data Augmentation		
Precision	Recall	F1-Score
.51	.49	.50
Multi-Feature DNN-50 Original Dataset		
Precision	Recall	F1-Score
.90	.91	.905

V. CONCLUSION

In this paper, we describe an algorithm for traffic classification in urban areas from acoustic data. Given the ubiquity of streetlights in urban areas, acoustic sensors mounted in smart streetlights can be used to capture surrounding noise generated by traffic as well as other sources such as passersby, wind gusts, rain, etc. However, some of the random noise can be filtered from the prevalent noise generated from the traffic. Subsequently, we identified the number of unique sources of acoustic signals using FFT for signal analysis with extracting the individual components of the audio signals using ICA. We also proposed a multi-feature generation methodology for our captured acoustics. The key contribution of this research is to demonstrate that multi-features along with transfer learning with cross-domain methodology are more promising in improving the accuracy of the classifiers.

Being able to classify traffic in urban areas can be a great tool for urban planning in terms of locating congestion areas, the prevalence of different types of vehicles, and diurnal patterns of traffic. Acoustic signals also provide better privacy than video capture. The lighter size of acoustic signals is another important advantage. In our future work, we are planning to develop a deep learning algorithm with more interpretability in working to solve other real-time problems like road traffic congestion detection and anomaly detection on road to create the surveillance system.

ACKNOWLEDGMENT

This work is supported by NSF GRANT IIS-1923982

REFERENCES

[1] S. Kozhisseri and M. Bikdash. Spectral features for the classification of civilian vehicles using acoustic sensors. In 2009IEEE Workshop on Computational Intelligence in Vehicles and vehicular Systems, pages 93–100, March 2009.

[2] Ali Dalir, Ali Asghar Beheshti Shirazi, and Morteza Hoseini Masoom. Classification of vehicles based on audio signals using quadratic discriminant analysis and high energy feature vectors. CoRR, abs/1804.01212, 2018.

[3] V. Ovchinnikov A. Grakovski. The analysis of possibility of acoustic sensors application for moving road vehicles detecting. In Proceedings of the 9th International Conference "Reliability and Statistics in Transportation and Communication" (Rel-Stat'09), 2009.

[4] Peter E. William and Michael W. Hoffman. Efficient sensor network vehicle classification using peak harmonics of acoustic emissions. In Edward M.Carapezza, editor, Unattended Ground, Sea, and Air Sensor Technologies and Applications X, volume 6963, pages 198 – 209. International Society for Optics and photonics, SPIE, 2008.

[5] Shigemi Ishida, Song Liu, Kohei Mimura, Shigeaki Tagashira, and Akira Fukuda. Design of acoustic vehicle count system using dtw. In ITS World Congress, Melbourne, Australia, 10 2016.

[6] Rijurekha Sen, Abhinav Maurya, Bhaskaran Raman, Rupesh Mehta, Ramakrishnan Kalyanaraman, Nagamanoj Vankadhara, Swaroop Roy, and Prashima Sharma. Kyun queue: A sensor network system to monitor road traffic queues. pages 127–140,11 2012.

[7] Maria Nadia Postorino and Giuseppe M. L. Sarn`e. An agent-based sensor grid to monitor urban traffic. CEUR Workshop Proceedings, 1260, 01 2014.

[8] M. V. Peppa, D. Bell, T. Komar, and W. Xiao. Urban traffic flow analysis based on deep learning car detection from cctv image series. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4:499–506,2018.

[9] Mohammadhani Fouladgar, Mostafa Parchami, Ramez Elmasri, and Amir Ghaderi. Scalable deep traffic flow neural networks for urban traffic congestion prediction. CoRR, abs/1703.01006,2017.

[10] Manuel Lopez-Martin, Bel´en Carro, Antonio Sanchez-Esguevillas, and Jaime Lloret. Network traffic classifier with convolutional and recurrent neural networks for internet of things. IEEE Access, PP:1–1, 09 2017.