Scalable Gunshot Detection Systems with Convolutional Neural Networks

1st Alex Morehead
Computer Science, Mathematics, & Physics
Missouri Western State University
Saint Joseph, Missouri, USA
amorehead1@missouriwestern.edu

2nd Lauren Ogden
Computer Science
Columbia University
New York City, New York, USA
lao2125@columbia.edu

3rd Gabe Magee
Computer Science
Pomona College
Claremont, California, USA
glma2016@pomona.edu

4th Ryan Hosler
Computer & Information Science
Indiana University-Purdue University Indianapolis
Indianapolis, Indiana, USA
rjhosler@iu.edu

5th George Mohler
Computer & Information Science
Indiana University-Purdue University Indianapolis
Indianapolis, Indiana, USA
gmohler@iupui.edu

***Abstract*—Many cities with gunshot detection systems depend on expensive systems that rely on humans differentiating between gunshots and non-gunshots, such as ShotSpotter®. Thus, a scalable gunshot detection system that is low in cost and high in accuracy would be advantageous for a variety of cities across the globe, in that it would favorably promote the delegation of tasks typically worked by humans to machines. A convolutional neural network (CNN) was trained on a variety of sound data to recognize gunshots. This model was then deployed to a Raspberry Pi Model 3 B+ with an SMS modem attached, and the results were found to be advantageous. The findings generated by this research project have the potential to expand the current state of knowledge regarding sound-based applications of CNNs, and while simultaneously reducing the amount of jobs that require human input the results of this project could very well lead to an increase in safety standards for a city's residents.**

***Index Terms*—machine learning, convolutional neural network, spectrogram, sound classification, microcomputer, smart city**

I. Introduction

Properly implementing a gunshot detection model to be used on a city-wide array of microcomputers enables automation of what previously required dedicated teams of human operators to perform. Further, it demonstrates the capabilities of deep learning architectures in sound classification from substantial amounts of sound data. Motivations such as these make it clear why this category of research is warranted along with our endeavors.

II. The Data

A. Sources and Derivatives

We obtained our data from two places: free internet databases such as Freesound and a repository of sounds recorded using a microphone connected to a Raspberry Pi microcomputer. In addition to this, we used a generative adversarial network (GAN) as well as sound augmentations to create additional samples of gunfire sounds and to prevent our model from overfitting to our compiled dataset.

Further, we decidedly divided our data into three sets: training data, testing data, and validation data. This was to prevent our models from overfitting to our given dataset which had been occurring previously. Detailed below is a list of the augmentations we applied to each of our sound samples.

TABLE I
Data Augmentations

Time Shift	Pitch Change	Speed Change
Shifts a sound sample to the left or right by a randomly chosen amount less than 50% of the length, and then fills in silence as needed.	Changes the pitch of a sample by a randomly-chosen factor between 70% and 130%.	Alters the playback speed of a sample by a randomly-chosen amount between 70% and 130%.

Volume Change	Background Noise Addition
Increases the amplitude of a sample with a uniformly-random variable.	Introduces random background noise into a sample while making sure that no gunshots are added into a sample that does not originally contain a gunshot.

Fig. 1. The data augmentations used on all sound samples.

III. CNNs

Convolutional Neural Networks (CNNs) are neural networks designed to locate, model, and accurately predict patterns

We gratefully acknowledge the support of NSF grant REU-1659488 which provided research stipends, travel funds, and supply money for our summer research project.

present in input data such as a colored image. They do so by iteratively sliding over small regions of data and translating any inherent properties in a region over to a proceeding network layer. This process is repeated up until the output layer which generates a prediction.

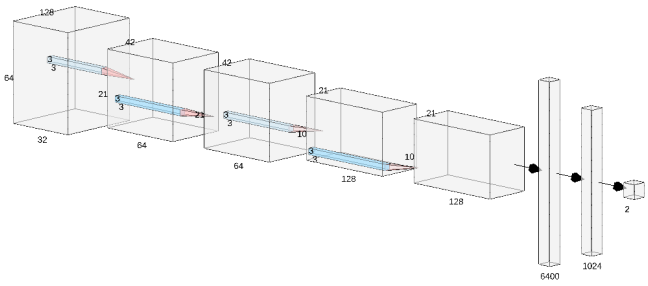

Fig. 2. An illustration of our 2D CNN's architecture.

IV. Spectrograms

Spectrograms are visual representations of the frequency and amplitude of sound over a specified span of time. For our project, we created two models, a 1D architecture that looks at sound represented as an array of frequency values and a 2D architecture that instead analyzes sound represented as spectrograms. For the time-series model each entry simply corresponds to a frequency measurement in a time-series, whereas for the spectrogram model the entry for each specific frequency-time cartesian coordinate is an amplitude value.

V. Methodology

A. Training

A convolutional neural network (CNN) was trained on a variety of sound data to recognize gunshots. While we had labels for sounds other than gunshots, we grouped them into a singular group "other". Then each of these samples were preprocessed to turn them into spectrograms if the model was trained on them. The models were trained for 100 epochs or until the target metric of a training session, accuracy in this case, did not change for fifteen epochs.

B. Deployment

Each model was then deployed to a Raspberry Pi Model 3 B+ with an SMS modem attached. In order to conserve space, we decided to convert them to TensorFlow Lite (TFLite) models, with a tradeoff for performance time. Our program has three processes – one to put audio from a stream onto a queue, one to analyze sound data pulled from the queue, and one to send an SMS alert message to a predetermined list of phone numbers.

VI. Results

Using a validation data set, we found that our Keras models performed reasonably well with predictions. Regarding the performance of our models when stored in TFLite format, the results below describe expected levels of effectiveness.

TABLE II
Gunshot Detection Results

	2D CNN (128 x 64)		**2D CNN (128 x 128)**		**CNN Ensemble**	
	Keras	*TFLite*	*Keras*	*TFLite*	*Keras*	*TFLite*
Accuracy	98.8%	N/A	98.8%	N/A	98.8%	N/A
Precision	96.5%	N/A	94.3%	N/A	97.1%	N/A
Recall	93.8%	N/A	96.2%	N/A	92.9%	N/A
F1 Score	95.1%	N/A	95.2%	N/A	95.0%	N/A

Fig. 3. The findings of a cross-evaluation technique applied to our models.

VII. Significance & Future Work

The findings generated by this research project have the potential to expand the current state of knowledge regarding sound-based applications of CNNs, and while simultaneously reducing the amount of jobs that require human input the results of this project could very well increase the standards of safety for a city's residents. Ideally, a feature we would like to implement in our pipeline in the future would allow for robust localization of gunshot alerts within a proposed cluster of Raspberry Pi units.

Acknowledgment

We gratefully acknowledge the support of NSF grant REU-1659488 which provided research stipends, travel funds, and supply money for this summer research project. Thanks also to all faculty and support staff who helped coordinate this year's undergraduate research experience (REU) for data science at IUPUI.

References

[1] Piczak, Karol J. "Environmental sound classification with convolutional neural networks." In 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1-6. IEEE, 2015.

[2] Takahashi, Naoya, Michael Gygli, Beat Pfister, and Luc Van Gool. "Deep convolutional neural networks and data augmentation for acoustic event detection." arXiv preprint arXiv:1604.07160 (2016).

[3] Takahashi, Naoya, Michael Gygli, Beat Pfister, and Luc Van Gool. "Deep convolutional neural networks and data augmentation for acoustic event detection." arXiv preprint arXiv:1604.07160 (2016).

[4] Lim, Hyungui, Jeongsoo Park, and Y. Han. "Rare sound event detection using 1D convolutional recurrent neural networks." In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017), pp. 80-84. 2017.

[5] Jaiswal, Kaustumbh, and Dhairya Kalpeshbhai Patel. "Sound Classification Using Convolutional Neural Networks." In 2018 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), pp. 81-84. IEEE, 2018.

[6] DAmiriparian, Shahin, N. Cummins, S. Julka, and B. W. Schuller. "Deep convolutional recurrent neural network for rare acoustic event detection." In Proc. DAGA, pp. 1522-1525. 2018.

[7] Shi, Bowen, Ming Sun, Chieh-Chi Kao, Viktor Rozgic, Spyros Matsoukas, and Chao Wang. "Compression of acoustic event detection models with low-rank matrix factorization and quantization training." arXiv preprint arXiv:1905.00855 (2019).

[8] Prince, Peter, Andrew Hill, Evelyn Piña Covarrubias, Patrick Doncaster, Jake L. Snaddon, and Alex Rogers. "Deploying acoustic detection algorithms on low-cost, open-source acoustic sensors for environmental monitoring." Sensors 19, no. 3 (2019): 553.