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[a].	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.

[a]Sample of a Table footnote.

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

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present

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

them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

III. CNNs

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

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. Gunshot Detection 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.

Fig. 2. Example of a figure caption.

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity "Magnetization", or "Magnetization, M", not just "M". If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write "Magnetization (A/m)" or "Magnetization {A[m(1)]}", not just "A/m". Do not label axes with a ratio of quantities and units. For example, write "Temperature (K)", not "Temperature/K".

TABLE II
Table Type Styles

Table Head	**Table Column Head**		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

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] Environmental Sound Classification with Convolutional Neural Networks, 2015
[2] Deep Convolutional Neural Networks and Data Augmentation for Environmental Sound Classification, 2016
[3] Deep Convolutional Neural Networks and Data Augmentation for Acoustic Event Detection, 2016
[4] Rare Sound Event Detection Using 1D Convolutional Recurrent Neural Networks, 2017
[5] Sound Classification Using Convolutional Neural Networks, 2018
[6] Deep Convolutional Recurrent Neural Network for Rare Acoustic Event Detection, 2018
[7] Compression of Acoustic Event Detection Models with Low-rank Matrix Factorization and Quantization Training, 2019
[8] Deploying Acoustic Detection Algorithms on Low-Cost, Open-Source Acoustic Sensors for Environmental Monitoring, 2019
[9]
[10] The following bibliography entries came with this LaTex template, and they can be used as a reference point for formatting and styling.
[11]
[12] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.
[13] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[14] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
[15] K. Elissa, “Title of paper if known,” unpublished.
[16] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
[17] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
[18] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.